ISSN 0090-5364 (print)
ISSN 2168-8966 (online)

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 53, No. 3—June 2025

THE ANNALS OF STATISTICS Vol. 53, No. 3, pp. 907–1369 June 2025

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

President: Tony Cai, Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104-6304, USA

President-Elect: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

Past President: Michael Kosorok, Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA

Executive Secretary: Peter Hoff, Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA

Treasurer: Jiashun Jin, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. *Editors*: Hans-Georg Müller, Department of Statistics, University of California, Davis, Davis, CA 95616, USA. Harrison Zhou, Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA

The Annals of Applied Statistics. *Editor-in-Chief*: Lexin Li, Department of Biostatistics and Epidemiology, University of California, Berkeley, Berkeley, CA 94720-7360, USA

The Annals of Probability. *Editors*: Paul Bourgade, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012-1185, USA. Julien Dubedat, Department of Mathematics, Columbia University, New York, NY 10027, USA

The Annals of Applied Probability. *Editors*: Jian Ding, School of Mathematical Sciences, Peking University, 100871, Bejing, China. Claudio Landim, IMPA, 22461-320, Rio de Janeiro, Brazil

Statistical Science. *Editor*: Moulinath Banerjee, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The IMS Bulletin. *Editor*: Tati Howell, bulletin@imstat.org

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 53, Number 3, June 2025. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, MD 21769, USA.

Copyright © 2025 by the Institute of Mathematical Statistics
Printed in the United States of America

The Annals of Statistics
2025, Vol. 53, No. 3, 907–928
https://doi.org/10.1214/24-AOS2462
© Institute of Mathematical Statistics, 2025

ASYMPTOTIC DISTRIBUTIONS OF LARGEST PEARSON CORRELATION COEFFICIENTS UNDER DEPENDENT STRUCTURES

BY TIEFENG JIANG[1,a] AND TUAN PHAM[2,b]

[1]*School of Data Science, Chinese University of Hong Kong,* [a]*jiang040@cuhk.edu.cn*
[2]*Department of Statistics and Data Science, University of Texas,* [b]*tuan.pham@utexas.edu*

Given a random sample from a multivariate normal distribution whose covariance matrix is a Toeplitz matrix, we study the largest off-diagonal entry of the sample correlation matrix. Assuming the multivariate normal distribution has the covariance structure of an autoregressive sequence, we establish a phase transition in the limiting distribution of the largest off-diagonal entry. We show that the limiting distributions are of Gumbel-type (with different parameters), depending on how large or small the parameter of the autoregressive sequence is. At the critical case, we obtain that the limiting distribution is the maximum of two independent random variables of Gumbel distributions. This phase transition establishes the exact threshold at which the autoregressive covariance structure behaves differently than its counterpart with the covariance matrix equal to the identity. Assuming the covariance matrix is a general Toeplitz matrix, we obtain the limiting distribution of the largest entry under the ultrahigh-dimensional settings: it is a weighted sum of two independent random variables, one normal and the other following a Gumbel-type law. The counterpart of the non-Gaussian case is also discussed. We use our results to obtain new testing procedures for some problems in high-dimensional statistics.

REFERENCES

BASTIAN, P., DETTE, H. and HEINY, J. (2024). Testing for practically significant dependencies in high dimensions via bootstrapping maxima of U-statistics. *Ann. Statist.* **52** 628–653. MR4744190 https://doi.org/10.1214/24-aos2361

BICKEL, P. J. and LEVINA, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

BICKEL, P. J. and LEVINA, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969 https://doi.org/10.1214/009053607000000758

CAI, T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annu. Rev. Stat. Appl.* **4** 423–446.

CAI, T., FAN, J. and JIANG, T. (2013). Distributions of angles in random packing on spheres. *J. Mach. Learn. Res.* **14** 1837–1864. MR3104497

CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949 https://doi.org/10.1198/jasa.2011.tm10560

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

CAI, T. T. and JIANG, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *J. Multivariate Anal.* **107** 24–39. MR2890430 https://doi.org/10.1016/j.jmva.2011.11.008

CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

MSC2020 subject classifications. Primary 62E20; secondary 62H15.
Key words and phrases. High-dimensional testing, extreme values, sample correlation matrix, uniform central limit theorem.

CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

CAI, T. T. and ZHANG, A. (2016). Inference for high-dimensional differential correlation matrices. *J. Multivariate Anal.* **143** 107–126. MR3431422 https://doi.org/10.1016/j.jmva.2015.08.019

CHEN, X. and LIU, W. (2018). Testing independence with high-dimensional correlated samples. *Ann. Statist.* **46** 866–894. MR3782387 https://doi.org/10.1214/17-AOS1571

CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. MR4583674 https://doi.org/10.1214/22-aap1870

CUTTING, C., PAINDAVEINE, D. and VERDEBOUT, T. (2017). Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives. *Ann. Statist.* **45** 1024–1058. MR3662447 https://doi.org/10.1214/16-AOS1473

CUTTING, C., PAINDAVEINE, D. and VERDEBOUT, T. (2022). Testing uniformity on high-dimensional spheres: The non-null behaviour of the Bingham test. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 567–602. MR4374686 https://doi.org/10.1214/21-aihp1168

FAN, J. and JIANG, T. (2019). Largest entries of sample correlation matrices from equi-correlated normal populations. *Ann. Probab.* **47** 3321–3374. MR4021253 https://doi.org/10.1214/19-AOP1341

FENG, L., JIANG, T., LIU, B. and XIONG, W. (2022). Max-sum tests for cross-sectional independence of high-dimensional panel data. *Ann. Statist.* **50** 1124–1143. MR4404930 https://doi.org/10.1214/21-aos2142

HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

HAN, Q., JIANG, T. and SHEN, Y. (2023). Contiguity under high-dimensional Gaussianity with applications to covariance testing. *Ann. Appl. Probab.* **33** 4272–4321. MR4674051 https://doi.org/10.1214/22-aap1917

HEINY, J. (2022). Large sample correlation matrices: A comparison theorem and its applications. *Electron. J. Probab.* **27** Paper No. 94, 20. MR4456777 https://doi.org/10.1214/22-ejp817

HEINY, J. and KLEEMANN, C. (2025). Maximum interpoint distance of high-dimensional random vectors. *Bernoulli* **31** 537–560. MR4815995 https://doi.org/10.3150/24-bej1738

HEINY, J. and MIKOSCH, T. (2018). Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices. *Stochastic Process. Appl.* **128** 2779–2815. MR3811704 https://doi.org/10.1016/j.spa.2017.10.002

HEINY, J., MIKOSCH, T. and YSLAS, J. (2021). Point process convergence for the off-diagonal entries of sample covariance matrices. *Ann. Appl. Probab.* **31** 538–560. MR4254488 https://doi.org/10.1214/20-aap1597

HEINY, J. and YAO, J. (2022). Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations. *Ann. Statist.* **50** 3249–3280. MR4524496 https://doi.org/10.1214/22-aos2226

ISHII, A., YATA, K. and AOSHIMA, M. (2021). Hypothesis tests for high-dimensional covariance structures. *Ann. Inst. Statist. Math.* **73** 599–622. MR4247072 https://doi.org/10.1007/s10463-020-00760-5

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

JIANG, T. (2019). Determinant of sample correlation matrix with application. *Ann. Appl. Probab.* **29** 1356–1397. MR3914547 https://doi.org/10.1214/17-AAP1362

JIANG, T. and PHAM, T. (2025). Supplement to "Asymptotic distributions of largest Pearson correlation coefficients under dependent structures." https://doi.org/10.1214/24-AOS2462SUPP

KATO, N., YAMADA, T. and FUJIKOSHI, Y. (2010). High-dimensional asymptotic expansion of LR statistic for testing intraclass correlation structure and its error bound. *J. Multivariate Anal.* **101** 101–112. MR2557621 https://doi.org/10.1016/j.jmva.2009.05.006

KOIKE, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. MR4273258 https://doi.org/10.1007/s42081-020-00096-7

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. *Springer Series in Statistics*. Springer, New York. MR0691492

LI, D., LIU, W.-D. and ROSALSKY, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148** 5–35. MR2653220 https://doi.org/10.1007/s00440-009-0220-z

LI, D., QI, Y. and ROSALSKY, A. (2012). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.* **111** 256–270. MR2944420 https://doi.org/10.1016/j.jmva.2012.04.002

LI, D. and ROSALSKY, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **16** 423–447. MR2209348 https://doi.org/10.1214/105051605000000773

LI, D. and XUE, L. (2015). Joint limiting laws for high-dimensional independence tests. Preprint. Available at arXiv:1512.08819.

LI, D., XUE, L. and ZOU, H. (2018). Applications of Peter Hall's martingale limit theory to estimating and testing high dimensional covariance matrices. *Statist. Sinica* **28** 2657–2670. MR3839878

LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539 https://doi.org/10.1214/08-AAP527

PAROLYA, N., HEINY, J. and KUROWICKA, D. (2024). Logarithmic law of large random correlation matrices. *Bernoulli* **30** 346–370. MR4665581 https://doi.org/10.3150/23-bej1600

QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40** 1285–1314. MR3015026 https://doi.org/10.1214/12-AOS1002

SASVÁRI, Z. (1998). On a classical theorem in the theory of Fourier integrals. *Proc. Amer. Math. Soc.* **126** 711–713. MR1469433 https://doi.org/10.1090/S0002-9939-98-04604-8

SHAO, Q.-M. and ZHOU, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42** 623–648. MR3178469 https://doi.org/10.1214/13-AOP837

SRIVASTAVA, M. S. and REID, N. (2012). Testing the structure of the covariance matrix with fewer observations than the dimension. *J. Multivariate Anal.* **112** 156–171. MR2957293 https://doi.org/10.1016/j.jmva.2012.06.004

SRIVASTAVA, M. S. and SINGULL, M. (2017). Testing sphericity and intraclass covariance structures under a growth curve model in high dimension. *Comm. Statist. Simulation Comput.* **46** 5740–5751. MR3698554 https://doi.org/10.1080/03610918.2016.1175623

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

YU, X., LI, D. and XUE, L. (2024). Fisher's combined probability test for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **119** 511–524. MR4713910 https://doi.org/10.1080/01621459.2022.2126781

YU, X., LI, D., XUE, L. and LI, R. (2023). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. *J. Amer. Statist. Assoc.* **118** 2548–2561. MR4681603 https://doi.org/10.1080/01621459.2022.2061354

ZHAO, K., ZOU, T., ZHENG, S. and CHEN, J. (2023). Hypothesis testing on compound symmetric structure of high-dimensional covariance matrix. *Comput. Statist. Data Anal.* **185** Paper No. 107779, 20. MR4593087 https://doi.org/10.1016/j.csda.2023.107779

ZHENG, S., CHEN, Z., CUI, H. and LI, R. (2019). Hypothesis testing on linear structures of high-dimensional covariance matrix. *Ann. Statist.* **47** 3300–3334. MR4025743 https://doi.org/10.1214/18-AOS1779

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. MR2327033 https://doi.org/10.1090/S0002-9947-07-04192-X

The Annals of Statistics
2025, Vol. 53, No. 3, 929–962
https://doi.org/10.1214/24-AOS2481
© Institute of Mathematical Statistics, 2025

ON THE CONVERGENCE OF COORDINATE ASCENT VARIATIONAL INFERENCE

BY ANIRBAN BHATTACHARYA[1,a], DEBDEEP PATI[2,b] AND YUN YANG[3,c]

[1]*Department of Statistics, Texas A&M University,* [a]*anirbanb@stat.tamu.edu*
[2]*Department of Statistics, University of Wisconsin-Madison,* [b]*dpati2@wisc.edu*
[3]*Department of Mathematics, University of Maryland at College Park,* [c]*yy84@umd.edu*

As a computational alternative to Markov chain Monte Carlo approaches, variational inference (VI) is becoming more and more popular for approximating intractable posterior distributions in large-scale Bayesian models due to its comparable efficacy and superior efficiency. Several recent works provide theoretical justifications of VI by proving its statistical optimality for parameter estimation under various settings; meanwhile, formal analysis on the algorithmic convergence aspects of VI is still largely lacking. In this paper, we consider the common coordinate ascent variational inference (CAVI) algorithm for implementing the mean-field (MF) VI towards optimizing a Kullback–Leibler divergence objective functional over the space of all factorized distributions. Focusing on the two-block case, we analyze the convergence of CAVI by leveraging the extensive toolbox from functional analysis and optimization. We provide general conditions for certifying global or local exponential convergence of CAVI. Specifically, a new notion of generalized correlation for characterizing the interaction between the constituting blocks in influencing the VI objective functional is introduced, which according to the theory, quantifies the algorithmic contraction rate of two-block CAVI. As illustrations, we apply the theory to a number of examples, and derive explicit problem-dependent upper bounds on the algorithmic contraction rate.

REFERENCES

[1] ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

[2] ALQUIER, P. and RIDGWAY, J. (2020). Concentration of tempered posteriors and of their variational approximations. *Ann. Statist.* **48** 1475–1497. MR4124331 https://doi.org/10.1214/19-AOS1855

[3] AMBROGIONI, L., GÜÇLÜ, U., GÜÇLÜTÜRK, Y., HINNE, M., VAN GERVEN, M. A. and MARIS, E. (2018). Wasserstein variational inference. *Adv. Neural Inf. Process. Syst.* **31**.

[4] ARNESE, M. and LACKER, D. (2024). Convergence of coordinate ascent variational inference for log-concave measures via optimal transport. arXiv preprint. Available at arXiv:2404.08792.

[5] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120. MR3611487 https://doi.org/10.1214/16-AOS1435

[6] BHATTACHARYA, A., PATI, D. and YANG, Y. (2025). Supplement to "On the Convergence of Coordinate Ascent Variational Inference." https://doi.org/10.1214/24-AOS2481SUPP

[7] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124

[8] BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

[9] BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

MSC2020 subject classifications. 62F15.
Key words and phrases. Functional optimization, Kullback–Leibler divergence, latent variable models, mean-field, sub-Gaussian, variational Bayes.

[10] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[11] CELENTANO, M., FAN, Z. and MEI, S. (2023). Local convexity of the TAP free energy and AMP convergence for $\mathbb{Z}_2$-synchronization. *Ann. Statist.* **51** 519–546. MR4600991 https://doi.org/10.1214/23-aos2257

[12] CORDERO-ERAUSQUIN, D. (2017). Transport inequalities for log-concave measures, quantitative forms, and applications. *Canad. J. Math.* **69** 481–501. MR3679684 https://doi.org/10.4153/CJM-2016-046-3

[13] DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

[14] GHORBANI, B., JAVADI, H. and MONTANARI, A. (2018). An instability in variational inference for topic models. In *International Conference on Machine Learning*.

[15] GHOSH, I., BHATTACHARYA, A. and PATI, D. (2022). Statistical optimality and stability of tangent transform algorithms in logit models. *J. Mach. Learn. Res.* **23** Paper No. [184], 42. MR4577137

[16] HALL, P., ORMEROD, J. T. and WAND, M. P. (2011). Theory of Gaussian variational approximation for a Poisson mixed model. *Statist. Sinica* **21** 369–389. MR2796867

[17] HALL, P., PHAM, T., WAND, M. P. and WANG, S. S. J. (2011). Asymptotic normality and valid inference for Gaussian variational approximation. *Ann. Statist.* **39** 2502–2532. MR2906876 https://doi.org/10.1214/11-AOS908

[18] HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926

[19] JIN, C., ZHANG, Y., BALAKRISHNAN, S., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). Local maxima in the likelihood of Gaussian mixture models: Structural results and algorithmic consequences. *Adv. Neural Inf. Process. Syst.* **29**.

[20] JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

[21] KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. MR0039968 https://doi.org/10.1214/aoms/1177729694

[22] LACKER, D., MUKHERJEE, S. and YEUNG, L. C. (2024). Mean field approximations via log-concavity. *Int. Math. Res. Not. IMRN* **7** 6008–6042. MR4728726 https://doi.org/10.1093/imrn/rnad302

[23] LATTIMORE, T. (2017). A scale free algorithm for stochastic bandits with bounded kurtosis. *Adv. Neural Inf. Process. Syst.* **30**.

[24] LI, Y. and TURNER, R. E. (2016). Rényi divergence variational inference. *Adv. Neural Inf. Process. Syst.* **29**.

[25] LUENBERGER, D. G. (1997). *Optimization by Vector Space Methods*. Wiley, New York.

[26] MACKAY, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge Univ. Press, New York. MR2012999

[27] MINKA, T. P. (2013). Expectation propagation for approximate Bayesian inference. arXiv preprint. Available at arXiv:1301.2294.

[28] MUKHERJEE, S. S., SARKAR, P., WANG, Y. R. and YAN, B. (2018). Mean field for the stochastic blockmodel: Optimization landscape and convergence issues. In *Advances in Neural Information Processing Systems* 10694–10704.

[29] NESTEROV, Y. (2018). Smooth convex optimization. In *Lectures on Convex Optimization* 59–137.

[30] ORMEROD, J. T. and WAND, M. P. (2010). Explaining variational approximations. *Amer. Statist.* **64** 140–153. MR2757005 https://doi.org/10.1198/tast.2010.09058

[31] ORMEROD, J. T. and WAND, M. P. (2012). Gaussian variational approximate inference for generalized linear mixed models. *J. Comput. Graph. Statist.* **21** 2–17. MR2913353 https://doi.org/10.1198/jcgs.2011.09118

[32] PARISI, G. (1988). *Statistical Field Theory. Frontiers in Physics*. Addison-Wesley, Reading.

[33] PATI, D., BHATTACHARYA, A. and YANG, Y. (2018). On statistical optimality of variational Bayes. In *International Conference on Artificial Intelligence and Statistics* 1579–1588. PMLR.

[34] PEYPOUQUET, J. (2015). *Convex Optimization in Normed Spaces*: *Theory, Methods and Examples*. *SpringerBriefs in Optimization*. Springer, Cham. MR3310025 https://doi.org/10.1007/978-3-319-13710-0

[35] PLUMMER, S., PATI, D. and BHATTACHARYA, A. (2020). Dynamics of coordinate ascent variational inference: A case study in 2D Ising models. *Entropy* **22** Paper No. 1263, 33. MR4222066 https://doi.org/10.3390/e22111263

[36] QUARTERONI, A., SACCO, R. and SALERI, F. (2006). Iterative methods for solving linear systems. In *Numerical Mathematics* 125–182.

[37] RAY, K. and SZABÓ, B. (2022). Variational Bayes for high-dimensional linear regression with sparse priors. *J. Amer. Statist. Assoc.* **117** 1270–1281. MR4480711 https://doi.org/10.1080/01621459.2020.1847121

[38] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

[39] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback-Leibler divergence. *IEEE Trans*. *Inf*. *Theory* **60** 3797–3820. MR3225930 https://doi.org/10.1109/TIT.2014.2320500

[40] WAINWRIGHT, M. J., JORDAN, M. I. et al. (2008). Graphical models, exponential families, and variational inference. *Found*. *Trends Mach*. *Learn*. **1** 1–305.

[41] WANG, B. and TITTERINGTON, D. M. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Anal*. **1** 625–650. MR2221291 https://doi.org/10.1214/06-BA121

[42] WANG, B. and TITTERINGTON, D. M. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Anal*. **1** 625–650. MR2221291 https://doi.org/10.1214/06-BA121

[43] WANG, Y. and BLEI, D. M. (2019). Frequentist consistency of variational Bayes. *J*. *Amer*. *Statist*. *Assoc*. **114** 1147–1161. MR4011769 https://doi.org/10.1080/01621459.2018.1473776

[44] WU, Y. and ZHOU, H. H. (2021). Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. *Math*. *Stat*. *Learn*. **4** 143–220. MR4383733 https://doi.org/10.4171/msl/29

[45] XU, J., HSU, D. J. and MALEKI, A. (2016). Global analysis of expectation maximization for mixtures of two Gaussians. *Adv*. *Neural Inf*. *Process*. *Syst*. **29**.

[46] YANG, Y., PATI, D. and BHATTACHARYA, A. (2020). α-variational inference with statistical guarantees. *Ann*. *Statist*. **48** 886–905. MR4102680 https://doi.org/10.1214/19-AOS1827

[47] ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *Ann*. *Statist*. **48** 2575–2598. MR4152113 https://doi.org/10.1214/19-AOS1898

[48] ZHANG, F. and GAO, C. (2020). Convergence rates of variational posterior distributions. *Ann*. *Statist*. **48** 2180–2207. MR4134791 https://doi.org/10.1214/19-AOS1883

The Annals of Statistics
2025, Vol. 53, No. 3, 963–988
https://doi.org/10.1214/24-AOS2482
© Institute of Mathematical Statistics, 2025

OPTIMAL TRANSPORT MAP ESTIMATION IN GENERAL FUNCTION SPACES

BY VINCENT DIVOL[1,a], JONATHAN NILES-WEED[2,3,b] AND ARAM-ALEXANDRE POOLADIAN[3,c]

[1]*CREST, ENSAE, IP Paris,* [a]*vincent.divol@ensae.fr*
[2]*Courant Institute of Mathematical Sciences, New York University,* [b]*jnw@cims.nyu.edu,* [c]*aram-alexandre.pooladian@nyu.edu*
[3]*Center for Data Science, New York University*

We study the problem of estimating a function T, given independent samples from a distribution P and from the pushforward distribution $T_\sharp P$. This setting is motivated by applications in the sciences, where T represents the evolution of a physical system over time, and in machine learning, where, for example, T may represent a transformation learned by a deep neural network trained for a generative modeling task. To ensure identifiability, we assume that $T = \nabla\varphi_0$ is the gradient of a convex function in which case T is known as an *optimal transport map*. Prior work has studied the estimation of T under the assumption that it lies in a Hölder class, but general theory is lacking. We present a unified methodology for obtaining rates of estimation of optimal transport maps in general function spaces. Our assumptions are significantly weaker than those appearing in the literature: we require only that the source measure P satisfy a Poincaré inequality and that the optimal map be the gradient of a smooth convex function that lies in a space whose metric entropy can be controlled. As a special case, we recover known estimation rates for Hölder transport maps but also obtain nearly sharp results in many settings not covered by prior work. For example, we provide the first statistical rates of estimation when P is the normal distribution and the transport map is given by an infinite-width shallow neural network.

REFERENCES

ADAMS, R. A. and FOURNIER, J. J. F. (2003). *Sobolev Spaces*, 2nd ed. *Pure and Applied Mathematics* (*Amsterdam*) **140**. Elsevier, Amsterdam. MR2424078

ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the* 34*th International Conference on Machine Learning*, *Vol*. 70 214–223.

BACH, F. (2017). Breaking the curse of dimensionality with convex neutral networks. *J. Mach. Learn. Res.* **18** Paper No. 19, 53. MR3634886

BAKRY, D., BARTHE, F., CATTIAUX, P. and GUILLIN, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** 60–66. MR2386063 https://doi.org/10.1214/ECP.v13-1352

BALABDAOUI, F., DOSS, C. R. and DUROT, C. (2021). Unlinked monotone regression. *J. Mach. Learn. Res.* **22** Paper No. 172, 60. MR4318528

BANDEIRA, A. S., PERRY, A. and WEIN, A. S. (2018). Notes on computational-to-statistical gaps: Predictions using statistical physics. *Port. Math.* **75** 159–186. MR3892753 https://doi.org/10.4171/PM/2014

BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

BECK, A. (2017). *First-Order Methods in Optimization*. *MOS-SIAM Series on Optimization* **25**. SIAM, Philadelphia, PA. MR3719240 https://doi.org/10.1137/1.9781611974997.ch1

BELOMESTNY, D., MOULINES, E., NAUMOV, A., PUCHKIN, N. and SAMSONOV, S. (2024). Rates of convergence for density estimation with generative adversarial networks. *J. Mach. Learn. Res.* **25** Paper No. [29], 47. MR4723879

MSC2020 subject classifications. 62G05.
Key words and phrases. Optimal transport, nonparametric estimation.

BELOMESTNY, D., NAUMOV, A., PUCHKIN, N. and SAMSONOV, S. (2023). Simultaneous approximation of a smooth function and its derivatives by deep neural networks with piecewise-polynomial activations. *Neural Netw.* **161** 242–253. https://doi.org/10.1016/j.neunet.2023.01.035

BERNSTEIN, D. J., BUCHMANN, J. and DAHMEN, W., eds. (2009). *Post-Quantum Cryptography*. Springer, Berlin. MR2590642 https://doi.org/10.1007/978-3-540-88702-7

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

BOBKOV, S. and LEDOUX, M. (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** 383–400. MR1440138 https://doi.org/10.1007/s004400050090

BOBKOV, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820

BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference on Learning Theory* 48–166. PMLR.

BRUNA, J., REGEV, O., SONG, M. J. and TANG, Y. (2021). Continuous LWE. In *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* 694–707. ACM, New York. MR4398874 https://doi.org/10.1145/3406325.3451000

BUBECK, S. (2015). Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.* **8** 231–357. https://doi.org/10.1561/2200000050

BUNEA, F., TSYBAKOV, A. B. and WEGKAMP, M. H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35** 1674–1697. MR2351101 https://doi.org/10.1214/009053606000001587

BUNNE, C., KRAUSE, A. and CUTURI, M. (2022). Supervised training of conditional Monge maps. arXiv preprint. Available at arXiv:2206.14262.

BUNNE, C., PAPAXANTHOS, L., KRAUSE, A. and CUTURI, M. (2022). Proximal optimal transport modeling of population dynamics. In *International Conference on Artificial Intelligence and Statistics* 6511–6528. PMLR.

CAFFARELLI, L. A. (2000). Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** 547–563. MR1800860 https://doi.org/10.1007/s002200000257

CARLIER, G., CHERNOZHUKOV, V. and GALICHON, A. (2016). Vector quantile regression: An optimal transport approach. *Ann. Statist.* **44** 1165–1192. MR3485957 https://doi.org/10.1214/15-AOS1401

CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450

CHEWI, S. and POOLADIAN, A.-A. (2023). An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *C. R. Math. Acad. Sci. Paris* **361** 1471–1482. MR4683324 https://doi.org/10.5802/crmath.486

DEB, N., GHOSAL, P. and SEN, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. arXiv preprint. Available at arXiv:2107.01718.

DEGROOT, M. H. and GOEL, P. K. (1980). Estimation of the correlation coefficient from a broken random sample. *Ann. Statist.* **8** 264–278. MR0560728

DEMETÇI, P., SANTORELLA, R., SANDSTEDE, B. and SINGH, R. (2022). Unsupervised integration of single-cell multi-omics datasets with disproportionate cell-type representation. In *Research in Computational Molecular Biology. Lecture Notes in Computer Science* **13278** 3–19. Springer, Cham. MR4485288 https://doi.org/10.1007/978-3-031-04749-7_1

DIVOL, V., NILES-WEED, J. and POOLADIAN, A.-A. (2025). Supplement to "Optimal transport map estimation in general function spaces." https://doi.org/10.1214/24-AOS2482SUPP

E, W., MA, C. and WU, L. (2022). The Barron space and the flow-induced function spaces for neural network models. *Constr. Approx.* **55** 369–406. MR4376565 https://doi.org/10.1007/s00365-021-09549-y

FEYDY, J., CHARLIER, B., VIALARD, F.-X. and PEYRÉ, G. (2017). Optimal transport for diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 291–299. Springer, Berlin.

FINLAY, C., GEROLIN, A., OBERMAN, A. M. and POOLADIAN, A.-A. (2020). Learning normalizing flows from Entropy-Kantorovich potentials. arXiv preprint. Available at arXiv:2006.06033.

FLAMARY, R., LOUNICI, K. and FERRARI, A. (2019). Concentration bounds for linear Monge mapping estimation and optimal transport domain adaptation. arXiv preprint. Available at arXiv:1905.10155.

GENEVAY, A., PEYRÉ, G. and CUTURI, M. (2018). Learning generative models with Sinkhorn divergences. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics* 1608–1617.

GINÉ, E. and NICKL, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics*, [40]. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

GOLDENSHLUGER, A. (2009). A universal procedure for aggregating estimators. *Ann. Statist.* **37** 542–568. MR2488362 https://doi.org/10.1214/00-AOS576

GOZLAN, N. (2010). Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 708–739. MR2682264 https://doi.org/10.1214/09-AIHP209

GRATHWOHL, W., CHEN, R. T., BETTENCOURT, J., SUTSKEVER, I. and DUVENAUD, D. (2018). Ffjord: Free-form continuous dynamics for scalable reversible generative models. arXiv preprint. Available at arXiv:1810.01367.

GUNSILIUS, F. and XU, Y. (2021). Matching for causal effects via multimarginal optimal transport. arXiv preprint. Available at arXiv:2112.04398.

HUANG, C.-W., CHEN, R. T. Q., TSIRIGOTIS, C. and COURVILLE, A. (2021). Convex potential flows: Universal probability distributions with optimal transport and convex optimization. In *International Conference on Learning Representations*.

HÜTTER, J.-C. and RIGOLLET, P. (2021). Minimax estimation of smooth optimal transport maps. *Ann. Statist.* **49** 1166–1194. MR4255123 https://doi.org/10.1214/20-aos1997

LECUÉ, G. and MENDELSON, S. (2009). Aggregation via empirical risk minimization. *Probab. Theory Related Fields* **145** 591–613. MR2529440 https://doi.org/10.1007/s00440-008-0180-8

LUCET, Y. (1997). Faster than the fast Legendre transform, the linear-time Legendre transform. *Numer. Algorithms* **16** 171–185. MR1619184 https://doi.org/10.1023/A:1019191114493

MAKKUVA, A., TAGHVAEI, A., OH, S. and LEE, J. (2020). Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning* 6672–6681. PMLR.

MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2024). Plugin estimation of smooth optimal transport maps. *Ann. Statist.* **52** 966–998. MR4784066 https://doi.org/10.1214/24-aos2379

MASSART, P. (2007). *Concentration Inequalities and Model Selection*: *Ecole d'Eté de Probabilités de Saint-Flour XXXIII*-2003. Springer, Berlin. MR2319879

MILGROM, P. and SEGAL, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica* **70** 583–601. MR1913824 https://doi.org/10.1111/1468-0262.00296

MORIEL, N., SENEL, E., FRIEDMAN, N., RAJEWSKY, N., KARAISKOS, N. and NITZAN, M. (2021). NovoSpaRc: Flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat. Protoc.* **16** 4177–4200. https://doi.org/10.1038/s41596-021-00573-7

MUZELLEC, B., VACHER, A., BACH, F., VIALARD, F.-X. and RUDI, A. (2021). Near-optimal estimation of smooth transport maps with kernel sums-of-squares. arXiv preprint. Available at arXiv:2112.01907.

NILES-WEED, J. and RIGOLLET, P. (2022). Estimation of Wasserstein distances in the spiked transport model. *Bernoulli* **28** 2663–2688. MR4474558 https://doi.org/10.3150/21-bej1433

PANANJADY, A., WAINWRIGHT, M. J. and COURTADE, T. A. (2016). Linear regression with an unknown permutation: Statistical and computational limits. In 2016 54*th Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton*) 417–424. IEEE.

POOLADIAN, A.-A. and NILES-WEED, J. (2021). Entropic estimation of optimal transport maps. arXiv preprint. Available at arXiv:2109.12004.

REGEV, O. (2005). On lattices, learning with errors, random linear codes, and cryptography. In *STOC'*05: *Proceedings of the* 37*th Annual ACM Symposium on Theory of Computing* 84–93. ACM, New York. MR2181605 https://doi.org/10.1145/1060590.1060603

REGEV, O. (2010). The learning with errors problem. In 25*th Annual IEEE Conference on Computational Complexity—CCC* 2010 191–204. IEEE Comput. Soc., Los Alamitos, CA. MR2932356

RIGOLLET, P. and TSYBAKOV, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. MR2356821 https://doi.org/10.3103/S1066530707030052

RIGOLLET, P. and WEED, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Inf. Inference* **8** 691–717. MR4045481 https://doi.org/10.1093/imaiai/iaz006

SALIMANS, T., ZHANG, H., RADFORD, A. and METAXAS, D. (2018). Improving GANs using optimal transport. In *International Conference on Learning Representations*.

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

SCHIEBINGER, G., SHU, J., TABAKA, M., CLEARY, B., SUBRAMANIAN, V., SOLOMON, A., GOULD, J., LIU, S., LIN, S. et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176** 928–943.

SCHÖLKOPF, B., SMOLA, A. J., BACH, F. et al. (2002). *Learning with Kernels*: *Support Vector Machines*, *Regularization*. *Optimization*, *and Beyond*. MIT press, Cambridge.

SLAWSKI, M. and SEN, B. (2024). Permuted and unlinked monotone regression in $\mathbb{R}^d$: An approach based on mixture modeling and optimal transport. *J. Mach. Learn. Res.* **25** Paper No. [183], 57. MR4777425

SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T. and GUIBAS, L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.* **34** 66.

SOLOMON, J., PEYRÉ, G., KIM, V. G. and SRA, S. (2016). Entropic metric alignment for correspondence problems. *ACM Trans. Graph.* **35** 72:1–72:13.

TALAGRAND, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes. Springer Monographs in Mathematics*. Springer, Berlin. MR2133757

TOROUS, W., GUNSILIUS, F. and RIGOLLET, P. (2021). An optimal transport approach to causal inference. arXiv preprint. Available at arXiv:2108.05858.

TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003. Proceedings* 303–313. Springer, Berlin.

TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

VACHER, A. and VIALARD, F.-X. (2021). Convex transport potential selection with semi-dual criterion. arXiv preprint. Available at arXiv:2112.07275.

VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

YANG, K. D., DAMODARAN, K., VENKATACHALAPATHY, S., SOYLEMEZOGLU, A. C., SHIVASHANKAR, G. and UHLER, C. (2020a). Predicting cell lineages using autoencoders and optimal transport. *PLoS Comput. Biol.* **16** e1007828.

YANG, Z., JIN, C., WANG, Z., WANG, M. and JORDAN, M. I. (2020b). On function approximation in reinforcement learning: Optimism in the face of large state spaces. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* 13903–13916.

ZHOU, D.-X. (2008). Derivative reproducing properties for kernel methods in learning theory. *J. Comput. Appl. Math.* **220** 456–463. MR2444183 https://doi.org/10.1016/j.cam.2007.08.023

The Annals of Statistics
2025, Vol. 53, No. 3, 989–1014
https://doi.org/10.1214/24-AOS2485
© Institute of Mathematical Statistics, 2025

SEMIPARAMETRIC INFERENCE BASED ON ADAPTIVELY COLLECTED DATA

BY LICONG LIN[1,a], KOULIK KHAMARU[2,b] AND MARTIN J. WAINWRIGHT[3,c]

[1]*Department of Statistics, University of California, Berkeley,* [a]*liconglin@berkeley.edu*

[2]*Department of Statistics, Rutgers University,* [b]*kk1241@stat.rutgers.edu*

[3]*Departments of Electrical Engineering and Computer Science and Mathematics, Massachusetts Institute of Technology,* [c]*wainwrigwork@gmail.com*

Many standard estimators, when applied to adaptively collected data, fail to be asymptotically normal, thereby complicating the construction of confidence intervals. We address this challenge in a semiparametric context: estimating the parameter vector of a generalized linear regression model contaminated by a nonparametric nuisance component. We construct suitably weighted estimating equations that account for adaptivity in data collection and provide conditions under which the associated estimates are asymptotically normal. Our results characterize the degree of "explorability" required for asymptotic normality to hold. For the simpler problem of estimating a linear functional, we provide similar guarantees under much weaker assumptions. We illustrate our general theory with concrete consequences for various problems, including standard linear bandits and sparse generalized bandits, and compare with other methods via simulation studies.

REFERENCES

[1] ABBASI-YADKORI, Y., PÁL, D. and SZEPESVÁRI, C. (2011). Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems* **24**.

[2] AGARWAL, A., HSU, D., KALE, S., LANGFORD, J., LI, L. and SCHAPIRE, R. (2014). Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning* 1638–1646. PMLR.

[3] AI, C. and CHEN, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71** 1795–1843. MR2015420 https://doi.org/10.1111/1468-0262.00470

[4] ANDREWS, D. W. K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62** 43–72. MR1258665 https://doi.org/10.2307/2951475

[5] ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709

[6] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2014). Pivotal estimation via square-root Lasso in nonparametric regression. *Ann. Statist.* **42** 757–788. MR3210986 https://doi.org/10.1214/14-AOS1204

[7] BELLONI, A., CHERNOZHUKOV, V. and WEI, Y. (2016). Post-selection inference for generalized linear models with many controls. *J. Bus. Econom. Statist.* **34** 606–619. MR3547999 https://doi.org/10.1080/07350015.2016.1166116

[8] BIBAUT, A., CHAMBAZ, A., DIMAKOPOULOU, M., KALLUS, N. and VAN DER LAAN, M. (2021). Post-contextual-bandit inference. *Adv. Neural Inf. Process. Syst.* **34** 28548–28559.

[9] BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671. MR0663424

[10] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[11] BODORY, H., HUBER, M. and LAFFÉRS, L. (2022). Evaluating (weighted) dynamic treatment effects by double machine learning. *Econom. J.* **25** 628–648. MR4530103 https://doi.org/10.1093/ectj/utac018

[12] BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. and LJUNG, G. M. (2016). *Time Series Analysis*: *Forecasting and Control*, 5th ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3379415

MSC2020 subject classifications. Primary 62E20; secondary 6208.

Key words and phrases. Semiparametric inference, two-stage weighted Z-estimator, adaptive data collection, generalized linear model.

[13] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[14] CHERNOZHUKOV, V., NEWEY, W., SINGH, R. and SYRGKANIS, V. (2022). Automatic debiased machine learning for dynamic treatment effects and general nested functionals. Preprint. Available at arXiv:2203.13887. https://doi.org/10.48550/ARXIV.2203.13887

[15] COLANGELO, K. and LEE, Y.-Y. (2020). Double debiased machine learning nonparametric inference with continuous treatments. Preprint. Available at arXiv:2004.03036.

[16] DEL MORAL, P. and NICLAS, A. (2018). A Taylor expansion of the square root matrix function. *J. Math. Anal. Appl.* **465** 259–266. MR3806701 https://doi.org/10.1016/j.jmaa.2018.05.005

[17] DESHPANDE, Y., MACKEY, L., SYRGKANIS, V. and TADDY, M. (2018). Accurate inference for adaptive linear models. In *International Conference on Machine Learning* 1194–1203. PMLR.

[18] DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431. MR0548036

[19] FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839 https://doi.org/10.1111/j.1467-9868.2011.01005.x

[20] FAN, Q., HSU, Y.-C., LIELI, R. P. and ZHANG, Y. (2022). Estimation of conditional average treatment effects with high-dimensional data. *J. Bus. Econom. Statist.* **40** 313–327. MR4356575 https://doi.org/10.1080/07350015.2020.1811102

[21] FIGUEROA, C. A., AGUILERA, A., CHAKRABORTY, B., MODIRI, A., AGGARWAL, J., DELIU, N., SARKAR, U., WILLIAMS, J. J. and LYLES, C. R. (2021). Adaptive learning algorithms to optimize mobile applications for behavioral health: Guidelines for design decisions. *J. Amer. Med. Inform. Assoc.* **28** 1225–1234. https://doi.org/10.1093/jamia/ocab001

[22] FONTAINE, X., PERRAULT, P., VALKO, M. and PERCHET, V. (2021). Online a-optimal design and active linear regression. In *International Conference on Machine Learning* 3374–3383. PMLR.

[23] FOSTER, D. J. and SYRGKANIS, V. (2023). Orthogonal statistical learning. *Ann. Statist.* **51** 879–908. MR4630373 https://doi.org/10.1214/23-AOS2258

[24] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. *Springer Series in Statistics*. Springer, New York. MR1920390 https://doi.org/10.1007/b97848

[25] HADAD, V., HIRSHBERG, D. A., ZHAN, R., WAGER, S. and ATHEY, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proc. Natl. Acad. Sci. USA* **118** Paper No. e2014602118, 10. MR4294058 https://doi.org/10.1073/pnas.2014602118

[26] HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. MR1612242 https://doi.org/10.2307/2998560

[27] HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2978290

[28] HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

[29] JAMIESON, K., MALLOY, M., NOWAK, R. and BUBECK, S. (2014). lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory* 423–439. PMLR.

[30] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

[31] KATO, M., MCALINN, K. and YASUI, S. (2021). The adaptive doubly robust estimator and a paradox concerning logging policy. *Adv. Neural Inf. Process. Syst.* **34** 1351–1364.

[32] KAUFMANN, E., CAPPÉ, O. and GARIVIER, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *J. Mach. Learn. Res.* **17** Paper No. 1, 42. MR3482921

[33] KENNEDY, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.* **17** 3008–3049. MR4667730 https://doi.org/10.1214/23-ejs2157

[34] KHAMARU, K., DESHPANDE, Y., MACKEY, L. and WAINWRIGHT, M. J. (2021). Near-optimal inference in adaptive linear regression. Preprint. Available at arXiv:2107.02266.

[35] KNAUS, M. C. (2022). Double machine learning-based programme evaluation under unconfoundedness. *Econom. J.* **25** 602–627. MR4530102 https://doi.org/10.1093/ectj/utac015

[36] KOSHY, T. (2009). *Catalan Numbers with Applications*. Oxford Univ. Press, Oxford. MR2526440

[37] LAI, T. L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Ann. Statist.* **22** 1917–1930. MR1329175 https://doi.org/10.1214/aos/1176325764

[38] LAI, T. L. and WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166. MR0642726

[39] LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

[40] LEWIS, G. and SYRGKANIS, V. (2021). Double/debiased machine learning for dynamic treatment effects. In *In Advances in Neural Information Processing Systems* **34** 22695–22707. Curran Associates, Red Hook.

[41] LIN, L., KHAMARU, K. and WAINWRIGHT, M. J (2025). Supplement to "Semiparametric inference based on adaptively collected data." https://doi.org/10.1214/24-AOS2485SUPP

[42] LIU, M., ZHANG, Y. and ZHOU, D. (2021). Double/debiased machine learning for logistic partially linear model. *Econom. J.* **24** 559–588. MR4310827 https://doi.org/10.1093/ectj/utab019

[43] MANNOR, S. and TSITSIKLIS, J. N. (2003/04). The sample complexity of exploration in the multi-armed bandit problem. *J. Mach. Learn. Res.* **5** 623–648. MR2247994

[44] NEWEY, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5** 99–135.

[45] NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. MR1303237 https://doi.org/10.2307/2951752

[46] NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Harald Cramér Volume* (U. Grenander, ed.) 213–234. Almqvist & Wiksell, Stockholm. MR0112201

[47] NIE, X., TIAN, X., TAYLOR, J. and ZOU, J. (2018). Why adaptively collected data have negative bias and how to correct for it. *Adv. Neural Inf. Process. Syst.* **84** 1261–1269.

[48] NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

[49] OH, M.-H., IYENGAR, G. and ZEEVI, A. (2021). Sparsity-agnostic lasso bandit. In *International Conference on Machine Learning* 8271–8280. PMLR.

[50] OPRESCU, M., SYRGKANIS, V. and WU, Z. S. (2019). Orthogonal random forest for causal inference. In *International Conference on Machine Learning* 4932–4941. PMLR.

[51] PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. *Lecture Notes in Statistics* **13**. Springer, New York. MR0675954

[52] ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129. MR1325119

[53] ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

[54] SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151. MR0856811 https://doi.org/10.1214/aos/1176350055

[55] SEMENOVA, V. and CHERNOZHUKOV, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *Econom. J.* **24** 264–289. MR4281225 https://doi.org/10.1093/ectj/utaa027

[56] SYRGKANIS, V. and ZHAN, R. (2023). Post-Episodic Reinforcement Learning Inference. https://doi.org/10.48550/ARXIV.2302.08854

[57] TEWARI, A. and MURPHY, S. A. (2017). From ads to interventions: Contextual bandits in mobile health. In *Mobile Health: Sensors, Analytic Methods, and Applications* 495–517.

[58] TRELLA, A. L., ZHANG, K. W., NAHUM-SHANI, I., SHETTY, V., DOSHI-VELEZ, F. and MURPHY, S. A. (2022). Reward design for an online reinforcement learning algorithm supporting oral self-care. Preprint. Available at arXiv:2208.07406.

[59] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[60] VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. *Springer Series in Statistics*. Springer, New York. MR2867111 https://doi.org/10.1007/978-1-4419-9782-1

[61] WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839

[62] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

[63] WANG, S., SPORREL, K., VAN HOOF, H., SIMONS, M., DE BOER, R. D. D., ETTEMA, D., NIBBELING, N., DEUTEKOM, M. and KRÖSE, B. (2021). Reinforcement learning to send reminders at right moments in smartphone exercise application: A feasibility study. *Int. J. Environ. Res. Public Health* **18**. https://doi.org/10.3390/ijerph18116059

[64] WHITE, J. S. (1959). The limiting distribution of the serial correlation coefficient in the explosive case. II. *Ann. Math. Statist.* **30** 831–834. MR0107348 https://doi.org/10.1214/aoms/1177706213

[65] YOM-TOV, E., FERARU, G., KOZDOBA, M., MANNOR, S., TENNENHOLTZ, M. and HOCHBERG, I. (2017). Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system. *J. Med. Internet Res.* **19** e338.

[66] ZHAN, R., HADAD, V., HIRSHBERG, D. A. and ATHEY, S. (2021). Off-policy evaluation via adaptive weighting with data from contextual bandits. In *Proceedings of the* 27*th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* 2125–2135.

[67] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

[68] ZHANG, K., JANSON, L. and MURPHY, S. (2021). Statistical inference with m-estimators on adaptively collected data. *Adv. Neural Inf. Process. Syst.* **34** 7460–7471.

[69] ZHANG, K., JANSON, L. and MURPHY, S. (2020). Inference for batched bandits. *Adv. Neural Inf. Process. Syst.* **33** 9818–9829.

The Annals of Statistics
2025, Vol. 53, No. 3, 1015–1043
https://doi.org/10.1214/24-AOS2487
© Institute of Mathematical Statistics, 2025

THE NUMERAIRE E-VARIABLE AND REVERSE INFORMATION PROJECTION

BY MARTIN LARSSON[1,a], AADITYA RAMDAS[2,b] AND JOHANNES RUF[3,c]

[1]*Department of Mathematical Sciences, Carnegie Mellon University,* [a]*larsson@cmu.edu*
[2]*Departments of Statistics and ML, Carnegie Mellon University,* [b]*aramdas@cmu.edu*
[3]*Department of Mathematics, London School of Economics,* [c]*j.ruf@lse.ac.uk*

We consider testing a composite null hypothesis $\mathcal{P}$ against a point alternative Q using e-variables, which are nonnegative random variables X such that $\mathbb{E}_{\mathsf{P}}[X] \le 1$ for every $\mathsf{P} \in \mathcal{P}$. This paper establishes a fundamental result: under no conditions whatsoever on $\mathcal{P}$ or Q, there exists a special e-variable X^* that we call the numeraire, which is strictly positive and satisfies $\mathbb{E}_{\mathsf{Q}}[X/X^*] \le 1$ for every other e-variable X. In particular, X^* is log-optimal in the sense that $\mathbb{E}_{\mathsf{Q}}[\log(X/X^*)] \le 0$. Moreover, X^* identifies a particular subprobability measure P^* via the density $d\mathsf{P}^*/d\mathsf{Q} = 1/X^*$. As a result, X^* can be seen as a generalized likelihood ratio of Q against $\mathcal{P}$. We show that P^* coincides with the reverse information projection (RIPr) when additional assumptions are made that are required for the latter to exist. Thus, P^* is a natural definition of the RIPr in the absence of any assumptions on $\mathcal{P}$ or Q. In addition to the abstract theory, we provide several tools for finding the numeraire and RIPr in concrete cases. We discuss several nonparametric examples where we can indeed identify the numeraire and RIPr, despite not having a reference measure. Our results have interpretations outside of testing in that they yield the optimal Kelly bet against $\mathcal{P}$ if we believe reality follows Q. We end with a more general optimality theory that goes beyond the ubiquitous logarithmic utility. We focus on certain power utilities, leading to reverse Rényi projections in place of the RIPr, which also always exist.

REFERENCES

AGRAWAL, S., JUNEJA, S. and GLYNN, P. (2020). Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory*. *Proc*. *Mach*. *Learn*. *Res*. (*PMLR*) **117** 50. Proceedings of Machine Learning Research PMLR. MR4165384

BARTL, D. and KUPPER, M. (2019). A pointwise bipolar theorem. *Proc*. *Amer*. *Math*. *Soc*. **147** 1483–1495. MR3910414 https://doi.org/10.1090/proc/14231

BECHERER, D. (2001). The numeraire portfolio for unbounded semimartingales. *Finance Stoch*. **5** 327–341. MR1849424 https://doi.org/10.1007/PL00013535

BRANNATH, W. and SCHACHERMAYER, W. (1999). A bipolar theorem for $L^0_+(\Omega, \mathcal{F}, \mathbb{P})$. In *Séminaire de Probabilités*, *XXXIII*. *Lecture Notes in Math*. **1709** 349–354. Springer, Berlin. MR1768009 https://doi.org/10.1007/BFb0096525

CASGRAIN, P., LARSSON, M. and ZIEGEL, J. (2024). Sequential testing for elicitable functionals via supermartingales. *Bernoulli* **30** 1347–1374. MR4699555 https://doi.org/10.3150/23-bej1634

COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley Interscience, Hoboken, NJ. MR2239987

CSISZÁR, I. and MATÚŠ, F. (2003). Information projections revisited. *IEEE Trans*. *Inf*. *Theory* **49** 1474–1490. MR1984936 https://doi.org/10.1109/TIT.2003.810633

CSISZÁR, I. and TUSNÁDY, G. (1984). Information geometry and alternating minimization procedures. *Statist*. *Decisions* 205–237. suppl. 1, Recent results in estimation theory and related topics. MR0785210

CVITANIĆ, J. and KARATZAS, I. (2001). Generalized Neyman–Pearson lemma via convex duality. *Bernoulli* **7** 79–97. MR1811745 https://doi.org/10.2307/3318603

MSC2020 subject classifications. Primary 62G10, 62F03; secondary 62L12, 46N10.
Key words and phrases. E-variables, log-optimality, reverse information projection, reverse Rényi projection, duality, bipolar.

DELBAEN, F. and SCHACHERMAYER, W. (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300** 463–520. MR1304434 https://doi.org/10.1007/BF01450498

EKELAND, I. and TÉMAM, R. (1999). *Convex Analysis and Variational Problems*, English ed. *Classics in Applied Mathematics* **28**. SIAM, Philadelphia, PA. Translated from the French. MR1727362 https://doi.org/10.1137/1.9781611971088

GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. M. (2024). Safe testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* with discussion.

HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **20** 225–241. MR0030730 https://doi.org/10.1214/aoms/1177730032

HONDA, J. and TAKEMURA, A. (2010). An asymptotically optimal bandit algorithm for bounded support models. In *Conference on Learning Theory* 67–79. Citeseer.

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probab. Surv.* **17** 257–317. MR4100718 https://doi.org/10.1214/18-PS321

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263. MR0356306

KARATZAS, I. and KARDARAS, C. (2007). The numéraire portfolio in semimartingale financial models. *Finance Stoch.* **11** 447–493. MR2335830 https://doi.org/10.1007/s00780-007-0047-3

KARATZAS, I. and KARDARAS, C. (2021). *Portfolio Theory and Arbitrage—a Course in Mathematical Finance*. *Graduate Studies in Mathematics* **214**. Amer. Math. Soc., Providence, RI. MR4422028 https://doi.org/10.1090/gsm/214

KARATZAS, I., LEHOCZKY, J. P., SHREVE, S. E. and XU, G.-L. (1991). Martingale and duality methods for utility maximization in an incomplete market. *SIAM J. Control Optim.* **29** 702–730. MR1089152 https://doi.org/10.1137/0329039

KONING, N. W. (2023). Post-hoc and anytime valid permutation and group invariance testing. arXiv preprint arXiv:2310.01153.

KRAMKOV, D. and SCHACHERMAYER, W. (1999). The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Ann. Appl. Probab.* **9** 904–950. MR1722287 https://doi.org/10.1214/aoap/1029962818

LARDY, T., GRÜNWALD, P. and HARREMOËS, P. (2024). Reverse information projections and optimal e-statistics. *IEEE Trans. Inf. Theory* **70** 7616–7631. MR4818350 https://doi.org/10.1109/tit.2024.3444458

LI, Q. J. (1999). Estimation of mixture models. Ph.D. thesis, Yale Univ.

LONG, J. B. (1990). The numeraire portfolio. *J. Financ. Econ.* **26** 29–69. https://doi.org/10.1016/0304-405X(90)90012-O

PLATEN, E. (2006). A benchmark approach to finance. *Math. Finance* **16** 131–151. MR2194899 https://doi.org/10.1111/j.1467-9965.2006.00265.x

RAMDAS, A., GRÜNWALD, P., VOVK, V. and SHAFER, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist. Sci.* **38** 576–601. MR4665027 https://doi.org/10.1214/23-sts894

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. arXiv:2009.03167.

ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41** 1397–1409. MR0277063 https://doi.org/10.1214/aoms/1177696786

SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J. R. Stat. Soc. A, General* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-aos2020

WASSERMAN, L., RAMDAS, A. and BALAKRISHNAN, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1–27. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

ZHANG, Z., RAMDAS, A. and WANG, R. (2024). On the existence of powerful p-values and e-values for composite hypotheses. *Ann. Statist.* **52** 2241–2267. MR4829487 https://doi.org/10.1214/24-aos2434

The Annals of Statistics
2025, Vol. 53, No. 3, 1044–1067
https://doi.org/10.1214/25-AOS2492
© Institute of Mathematical Statistics, 2025

A DUALITY FRAMEWORK FOR ANALYZING RANDOM FEATURE AND TWO-LAYER NEURAL NETWORKS

BY HONGRUI CHEN[1,a], JIHAO LONG[2,b] AND LEI WU[3,c]

[1]*Department of Mathematics, Stanford University,* [a]*hongrui@stanford.edu*
[2]*Institute for Advanced Algorithms Research, Shanghai,* [b]*longjh1998@gmail.com*
[3]*School of Mathematical Sciences, Peking University,* [c]*leiwu@math.pku.edu.cn*

We consider the problem of learning functions within the $\mathcal{F}_{p,\pi}$ and Barron spaces, which play crucial roles in understanding random feature models (RFMs), two-layer neural networks as well as kernel methods. Leveraging tools from information-based complexity (IBC), we establish a dual equivalence between approximation and estimation and then apply it to study the learning of the preceding function spaces. The duality allows us to focus on the more tractable problem between approximation and estimation. To showcase the efficacy of our duality framework, we delve into two important but under-explored problems:

(1) *Random feature learning beyond kernel regime.* We derive sharp bounds for learning $\mathcal{F}_{p,\pi}$ using RFMs. Notably, the learning is efficient without the curse of dimensionality for $p > 1$. This underscores the extended applicability of RFMs beyond the traditional kernel regime, since $\mathcal{F}_{p,\pi}$ with $p < 2$ is strictly larger than the corresponding reproducing kernel Hilbert space (RKHS) where $p = 2$.

(2) *The L^∞ learning of RKHS.* We establish sharp, spectrum-dependent characterizations for the convergence of L^∞ learning error in both noiseless and noisy settings. Surprisingly, we show that popular kernel ridge regression can achieve near-optimal performance in L^∞ learning, despite it primarily minimizing square loss.

To establish the aforementioned duality, we introduce a type of IBC, termed I-complexity, to measure the size of a function class. Notably, I-complexity offers a tight characterization of learning in noiseless settings, yields lower bounds comparable to Le Cam's in noisy settings, and is versatile in deriving upper bounds. We believe that our duality framework holds potential for broad application in learning analysis across more scenarios.

REFERENCES

[1] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. MR0051437 https://doi.org/10.2307/1990404

[2] ARORA, S., DU, S. S., HU, W., LI, Z., SALAKHUTDINOV, R. R. and WANG, R. (2019). On exact computation with an infinitely wide neural net. In *Advances in Neural Information Processing Systems* 8141–8150.

[3] BACH, F. (2017). On the equivalence between kernel quadrature rules and random feature expansions. *J. Mach. Learn. Res.* **18** 21. MR3634888

[4] BACH, F. (2017). Breaking the curse of dimensionality with convex neutral networks. *J. Mach. Learn. Res.* **18** 19. MR3634886

[5] BARRON, A. R. (1992). Neural net approximation. In *Proc. 7th Yale Workshop on Adaptive and Learning Systems* **1** 69–72.

[6] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

[7] BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.

MSC2020 subject classifications. 62G08.
Key words and phrases. Neural network, random feature model, kernel ridge regression, reproducing kernel Hilbert space, nonparametric regression, curse of dimensionality.

[8] BARTOLUCCI, F., DE VITO, E., ROSASCO, L. and VIGOGNA, S. (2023). Understanding neural networks with reproducing kernel Banach spaces. *Appl. Comput. Harmon. Anal.* **62** 194–236. MR4484791 https://doi.org/10.1016/j.acha.2022.08.006

[9] BELLMAN, R. (1966). Dynamic programming. *Science* **153** 34–37.

[10] BIETTI, A. and BACH, F. (2021). Deep equals shallow for ReLU networks in kernel regimes. In *International Conference on Learning Representations*.

[11] BREZIS, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext*. Springer, New York. MR2759829

[12] BURQ, N. and LEBEAU, G. (2014). Probabilistic Sobolev embeddings, applications to eigenfunctions estimates. In *Geometric and Spectral Analysis. Contemp. Math.* **630** 307–318. Amer. Math. Soc., Providence, RI. MR3328548 https://doi.org/10.1090/conm/630/12672

[13] CAO, Y. and GU, Q. (2019). Generalization bounds of stochastic gradient descent for wide and deep neural networks. *Adv. Neural Inf. Process. Syst.* **32**.

[14] CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. MR2335249 https://doi.org/10.1007/s10208-006-0196-8

[15] CARRATINO, L., RUDI, A. and ROSASCO, L. (2018). Learning with SGD and random features. In *Proceedings of the* 32*nd International Conference on Neural Information Processing Systems* 10213–10224.

[16] CELENTANO, M., MISIAKIEWICZ, T. and MONTANARI, A. (2021). Minimum complexity interpolation in random features models. arXiv preprint. Available at arXiv:2103.15996.

[17] CHEN, H., LONG, J. and WU, L. (2025). Supplement to "A duality framework for analyzing random feature and two-layer neural networks." https://doi.org/10.1214/25-AOS2492SUPP

[18] CHIZAT, L., OYALLON, E. and BACH, F. (2019). On lazy training in differentiable programming. *Adv. Neural Inf. Process. Syst.* **32**.

[19] CHO, Y. and SAUL, L. K. (2009). Kernel methods for deep learning. In *Proceedings of the* 22*nd International Conference on Neural Information Processing Systems* 342–350.

[20] CUI, H., LOUREIRO, B., KRZAKALA, F. and ZDEBOROVÁ, L. (2021). Generalization error rates in kernel regression: The crossover from the noiseless to noisy regime. *Adv. Neural Inf. Process. Syst.* **34** 10131–10143.

[21] DANIELY, A. (2017). SGD learns the conjugate kernel class of the network. In *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* 2419–2427.

[22] DEVORE, R. A. (1998). Nonlinear approximation. In *Acta Numerica*, 1998. *Acta Numer.* **7** 51–150. Cambridge Univ. Press, Cambridge. MR1689432 https://doi.org/10.1017/S0962492900002816

[23] DICKER, L. H., FOSTER, D. P. and HSU, D. (2017). Kernel ridge vs. principal component regression: Minimax bounds and the qualification of regularization operators. *Electron. J. Stat.* **11** 1022–1047. MR3629418 https://doi.org/10.1214/17-EJS1258

[24] DURRETT, R. (2019). *Probability—Theory and Examples. Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge Univ. Press, Cambridge. MR3930614 https://doi.org/10.1017/9781108591034

[25] E, W. (2021). The dawning of a new era in applied mathematics. *Notices Amer. Math. Soc.* **68** 565–571. MR4228132 https://doi.org/10.1090/noti

[26] E, W., MA, C. and WU, L. (2019). A priori estimates of the population risk for two-layer neural networks. *Commun. Math. Sci.* **17** 1407–1425. MR4044196 https://doi.org/10.4310/CMS.2019.v17.n5.a11

[27] E, W., MA, C. and WU, L. (2020). A comparative analysis of optimization and generalization properties of two-layer neural network and random feature models under gradient descent dynamics. *Sci. China Math.* **63** 1235–1258. MR4119555 https://doi.org/10.1007/s11425-019-1628-5

[28] E, W., MA, C. and WU, L. (2022). The Barron space and the flow-induced function spaces for neural network models. *Constr. Approx.* **55** 369–406. MR4376565 https://doi.org/10.1007/s00365-021-09549-y

[29] E, W. and WOJTOWYTSCH, S. (2022). Representation formulas and pointwise properties for Barron functions. *Calc. Var. Partial Differential Equations* **61** 46. MR4375792 https://doi.org/10.1007/s00526-021-02156-6

[30] E, W. and YU, B. (2018). The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat.* **6** 1–12. MR3767958 https://doi.org/10.1007/s40304-018-0127-z

[31] FISCHER, S. and STEINWART, I. (2020). Sobolev norm learning rates for regularized least-squares algorithms. *J. Mach. Learn. Res.* **21** 205. MR4209491

[32] HAN, J., JENTZEN, A. and E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA* **115** 8505–8510. MR3847747 https://doi.org/10.1073/pnas.1718942115

[33] HASHEMI, A., SCHAEFFER, H., SHI, R., TOPCU, U., TRAN, G. and WARD, R. (2023). Generalization bounds for sparse random feature expansions. *Appl. Comput. Harmon. Anal.* **62** 310–330. MR4493015 https://doi.org/10.1016/j.acha.2022.08.003

[34] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

[35] JACOT, A., GABRIEL, F. and HONGLER, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems* 8571–8580.

[36] KALAVASIS, A., ZADIK, I. and ZAMPETAKIS, M. (2024). Transfer learning beyond bounded density ratios. arXiv preprint. Available at arXiv:2403.11963.

[37] KUO, F. Y., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (2008). Multivariate L_∞ approximation in the worst case setting over reproducing kernel Hilbert spaces. *J. Approx. Theory* **152** 135–160. MR2422145 https://doi.org/10.1016/j.jat.2007.11.006

[38] KUO, F. Y., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (2009). On the power of standard information for multivariate approximation in the worst case setting. *J. Approx. Theory* **158** 97–125. MR2523723 https://doi.org/10.1016/j.jat.2008.01.011

[39] KURKOVÁ, V. and SANGUINETI, M. (2001). Bounds on rates of variable-basis and neural-network approximation. *IEEE Trans. Inf. Theory* **47** 2659–2665. MR1873958 https://doi.org/10.1109/18.945285

[40] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer, Berlin.

[41] LI, Y., ZHANG, H. and LIN, Q. (2024). On the asymptotic learning curves of kernel ridge regression under power-law decay. *Adv. Neural Inf. Process. Syst.* **36**.

[42] LIU, F., HUANG, X., CHEN, Y. and SUYKENS, J. A. (2021). Random features for kernel approximation: A survey on algorithms, theory, and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 7128–7148.

[43] LONG, J. and HAN, J. (2022). Perturbational complexity by distribution mismatch: A systematic analysis of reinforcement learning in reproducing kernel Hilbert space. *J. Mach. Learn.* **1** 1–37. MR4716372

[44] LONG, J. and HAN, J. (2023). Reinforcement learning with function approximation: From linear to nonlinear. *J. Mach. Learn.* **2** 161–193. MR4734710

[45] MA, C., PATHAK, R. and WAINWRIGHT, M. J. (2023). Optimally tackling covariate shift in RKHS-based nonparametric regression. *Ann. Statist.* **51** 738–761. MR4601000 https://doi.org/10.1214/23-aos2268

[46] MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2022). Generalization error of random feature and kernel methods: Hypercontractivity and kernel matrix concentration. *Appl. Comput. Harmon. Anal.* **59** 3–84. MR4412180 https://doi.org/10.1016/j.acha.2021.12.003

[47] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.* **75** 667–766. MR4400901 https://doi.org/10.1002/cpa.22008

[48] MERCER, J. (1909). Functions of positive and negative type, and their connection the theory of integral equations. *Philos. Trans. R. Soc. Lond.* **209** 415–446.

[49] MICCHELLI, C. A. and RIVLIN, T. J. (1977). A survey of optimal recovery. In *Optimal Estimation in Approximation Theory* (*Proc. Internat. Sympos., Freudenstadt*, 1976). *The IBM Research Symposia Series* 1–54. Plenum, New York. MR0617931

[50] NOVAK, E. (1988). *Deterministic and Stochastic Error Bounds in Numerical Analysis*. *Lecture Notes in Math.* **1349**. Springer, Berlin. MR0971255 https://doi.org/10.1007/BFb0079792

[51] NOVAK, E. (2001). Quantum complexity of integration. *J. Complexity* **17** 2–16. MR1817606 https://doi.org/10.1006/jcom.2000.0566

[52] NOVAK, E. and WOŹNIAKOWSKI, H. (2008). *Tractability of Multivariate Problems: Standard Information for Functionals* **2**. Eur. Math. Soc., Zurich.

[53] NOVAK, E. and WOŹNIAKOWSKI, H. (2012). *Tractability of Multivariate Problems. Volume III: Standard Information for Operators*. *EMS Tracts in Mathematics* **18**. European Mathematical Society (EMS), Zürich. MR2987170 https://doi.org/10.4171/116

[54] PILLAUD-VIVIEN, L., RUDI, A. and BACH, F. (2018). Statistical optimality of stochastic gradient descent on hard learning problems through multiple passes. *Adv. Neural Inf. Process. Syst.* **31**.

[55] POZHARSKA, K. and ULLRICH, T. (2022). A note on sampling recovery of multivariate functions in the uniform norm. *SIAM J. Numer. Anal.* **60** 1363–1384. MR4436257 https://doi.org/10.1137/21M1410580

[56] RAHIMI, A. and RECHT, B. (2007). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems* **20**.

[57] RAHIMI, A. and RECHT, B. (2008). Uniform approximation of functions with random bases. In 2008 46*th Annual Allerton Conference on Communication, Control, and Computing* 555–561. IEEE, Los Alamitos.

[58] Raissi, M., Perdikaris, P. and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378** 686–707. MR3881695 https://doi.org/10.1016/j.jcp.2018.10.045

[59] Ritter, K. (2000). *Average-Case Analysis of Numerical Problems*. *Lecture Notes in Math.* **1733**. Springer, Berlin. MR1763973 https://doi.org/10.1007/BFb0103934

[60] Scetbon, M. and Harchaoui, Z. (2021). A spectral analysis of dot-product kernels. In *International Conference on Artificial Intelligence and Statistics* 3394–3402. PMLR.

[61] Schölkopf, B., Herbrich, R. and Smola, A. J. (2001). A generalized representer theorem. In *Computational Learning Theory* (*Amsterdam*, 2001). *Lecture Notes in Computer Science* **2111** 416–426. Springer, Berlin. MR2042050 https://doi.org/10.1007/3-540-44581-1_27

[62] Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423. MR0026286 https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[63] Siegel, J. W. and Xu, J. (2023). Characterization of the variation spaces corresponding to shallow neural networks. *Constr. Approx.* **57** 1109–1132. MR4601174 https://doi.org/10.1007/s00365-023-09626-4

[64] Sirignano, J. and Spiliopoulos, K. (2018). DGM: A deep learning algorithm for solving partial differential equations. *J. Comput. Phys.* **375** 1339–1364. MR3874585 https://doi.org/10.1016/j.jcp.2018.08.029

[65] Smale, S. and Zhou, D.-X. (2007). Learning theory estimates via integral operators and their approximations. *Constr. Approx.* **26** 153–172. MR2327597 https://doi.org/10.1007/s00365-006-0659-y

[66] Spek, L., Heeringa, T. J. and Brune, C. (2022). Duality for neural networks through reproducing kernel Banach spaces. arXiv preprint. Available at arXiv:2211.05020.

[67] Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer, Berlin.

[68] Steinwart, I., Hush, D. R. and Scovel, C. (2009). Optimal rates for regularized least squares regression. In *Conference on Learning Theory* 79–93.

[69] Steinwart, I. and Scovel, C. (2012). Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35** 363–417. MR2914365 https://doi.org/10.1007/s00365-012-9153-3

[70] Traub, J. F. (1982). *Iterative Methods for the Solution of Equations* **312**. Amer. Math. Soc., Providence.

[71] Traub, J. F. (2003). Information-based complexity. In *Encyclopedia of Computer Science* 850–854.

[72] Traub, J. F. (2009). A brief history of information-based complexity. In *Essays on the Complexity of Continuous Problems* 61–71. Eur. Math. Soc., Zürich. MR2657123 https://doi.org/10.4171/069-1/3

[73] Wahba, G. (1990). *Spline Models for Observational Data*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

[74] Wainwright, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

[75] Werschulz, A. G. (1991). *The Computational Complexity of Differential and Integral Equations*: *An Information-Based Approach*. Oxford Univ. Press, Oxford.

[76] Willem, M. (2012). *Minimax Theorems* **24**. Springer, Berlin.

[77] Woźniakowski, H. (1991). Average case complexity of multivariate integration. *Bull. Amer. Math. Soc.* (*N.S.*) **24** 185–194. MR1072015 https://doi.org/10.1090/S0273-0979-1991-15985-9

[78] Wu, L. and Long, J. (2022). A spectral-based analysis of the separation between two-layer neural networks and linear methods. *J. Mach. Learn. Res.* **23** 119. MR4576704

[79] Xie, Y., Shi, R., Schaeffer, H. and Ward, R. (2022). Shrimp: Sparser random feature models via iterative magnitude pruning. In *Mathematical and Scientific Machine Learning* 303–318. PMLR.

[80] Yang, Y. and Barron, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142

[81] Zhang, H., Xu, Y. and Zhang, J. (2009). Reproducing kernel Banach spaces for machine learning. *J. Mach. Learn. Res.* **10** 2741–2775. MR2579912

The Annals of Statistics
2025, Vol. 53, No. 3, 1068–1094
https://doi.org/10.1214/25-AOS2493
© Institute of Mathematical Statistics, 2025

BELIEF IN DEPENDENCE: LEVERAGING ATOMIC LINEARITY IN DATA BITS FOR RETHINKING GENERALIZED LINEAR MODELS

BY BENJAMIN BROWN[1,a], KAI ZHANG[1,b] AND XIAO-LI MENG[2,c]

[1]*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill,* [a]*brownb1@live.unc.edu,* [b]*zhangk@email.unc.edu*

[2]*Department of Statistics, Harvard University,* [c]*meng@stat.harvard.edu*

Two linearly uncorrelated binary variables must be also independent because nonlinear dependence cannot manifest with only two possible states. This inherent linearity is the atom of dependency constituting any complex form of relationship. Inspired by this observation, we develop a framework called binary expansion linear effect (BELIEF) for understanding arbitrary relationships with a binary outcome. Models from the BELIEF framework are easily interpretable because they describe the association of binary variables in the language of linear models, yielding convenient theoretical insight and striking Gaussian parallels. With BELIEF, one may study generalized linear models (GLM) through transparent linear models, providing insight into how the choice of link affects modeling. For example, setting a GLM interaction coefficient to zero does not necessarily lead to the kind of no-interaction model assumption as understood under their linear model counterparts. Furthermore, for a binary response, maximum likelihood estimation for GLMs paradoxically fails under complete separation, when the data are most discriminative, whereas BELIEF estimation automatically reveals the perfect predictor in the data that is responsible for complete separation. We explore these phenomena and provide related theoretical results. We also provide preliminary empirical demonstration of some theoretical results.

REFERENCES

[1] AGRESTI, A. and KATERI, M. (2011). *Categorical Data Analysis*. Springer.
[2] ALBERT, A. and ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10. MR0738319 https://doi.org/10.1093/biomet/71.1.1
[3] ANGRIST, J. D. and PISCHKE, J.-S. (2009). *Mostly Harmless Econometrics*: *An Empiricist's Companion*. Princeton Univ. Press.
[4] ANTONIADIS, A. and OPPENHEIM, G. (2012). *Wavelets and Statistics*. Springer.
[5] BAHADUR, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction*. *Stanford Mathematical Studies in the Social Sciences*, *VI* 158–168. Stanford Univ. Press, Stanford, CA. MR0121893
[6] BARBER, R. F. (2020). Is distribution-free inference possible for binary regression? *Electron*. *J*. *Stat*. **14** 3487–3524. MR4154845 https://doi.org/10.1214/20-EJS1749
[7] BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). The limits of distribution-free conditional predictive inference. *Inf*. *Inference* **10** 455–482. MR4270755 https://doi.org/10.1093/imaiai/iaaa017
[8] BERRY, W. D., DEMERITT, J. H. and ESAREY, J. (2010). Testing for interaction in binary logit and probit models: Is a product term essential? *Amer*. *J*. *Polit*. *Sci*. **54** 248–266.
[9] BROWN, B., ZHANG, K. and MENG, X.-L. (2025). Supplement to "BELIEF in Dependence: Leveraging Atomic Linearity in Data Bits for Rethinking Generalized Linear Models." https://doi.org/10.1214/25-AOS2493SUPPA, https://doi.org/10.1214/25-AOS2493SUPPB
[10] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*: *Methods*, *Theory and Applications*. *Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

MSC2020 subject classifications. Primary 62G05; secondary 62J05.
Key words and phrases. Binary expansion, distribution-free, multiresolution models, nonparametric statistics.

[11] BUJA, A., BROWN, L., BERK, R., GEORGE, E., PITKIN, E., TRASKIN, M., ZHANG, K. and ZHAO, L. (2019). Models as approximations I: Consequences illustrated with linear regression. *Statist. Sci.* **34** 523–544. MR4048582 https://doi.org/10.1214/18-STS693

[12] CHEN, Y., LI, X., LIU, J. and YING, Z. (2016). A fused latent and graphical model for multivariate binary data. arXiv preprint arXiv:1606.08925.

[13] CHERNOFF, H. (2004). Information, for testing the equality of two probabilities, from the margins of the 2×2 table. *J. Statist. Plann. Inference* **121** 209–214. MR2038817 https://doi.org/10.1016/S0378-3758(03)00135-6

[14] CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2023). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.* **10** 427–449. MR4567800 https://doi.org/10.1146/annurev-statistics-040120-022239

[15] COX, D. R. (1972). The analysis of multivariate binary data. *Appl. Stat.* 113–120.

[16] DAI, B., DING, S. and WAHBA, G. (2013). Multivariate Bernoulli distribution. *Bernoulli* **19** 1465–1483. MR3102559 https://doi.org/10.3150/12-BEJSP10

[17] ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466–1474. MR0217826 https://doi.org/10.1214/aoms/1177698701

[18] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

[19] FAN, Y., KONG, Y., LI, D. and ZHENG, Z. (2015). Innovated interaction screening for high-dimensional nonlinear classification. *Ann. Statist.* **43** 1243–1272. MR3346702 https://doi.org/10.1214/14-AOS1308

[20] GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (1995). *Bayesian Data Analysis. Texts in Statistical Science Series*. CRC Press, London. MR1385925

[21] GELMAN, A. and HENNIG, C. (2017). Beyond subjective and objective in statistics. *J. R. Stat. Soc. A, General* **180** 967–1033. MR3723784 https://doi.org/10.1111/rssa.12276

[22] GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

[23] GUPTA, C., PODKOPAEV, A. and RAMDAS, A. (2020). Distribution-free binary classification: Prediction sets, confidence intervals and calibration. *Adv. Neural Inf. Process. Syst.* **33** 3711–3723.

[24] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

[25] HE, X., WANG, L. and HONG, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41** 342–369. MR3059421 https://doi.org/10.1214/13-AOS1087

[26] LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications. MR1419991

[27] LEI, J., G'SELL, M., RINALDO, A., TIBSHIRANI, R. J. and WASSERMAN, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342 https://doi.org/10.1080/01621459.2017.1307116

[28] LI, J., ZHANG, W., WANG, P., LI, Q., ZHANG, K. and LIU, Y. (2023). Nonparametric prediction distribution from resolution-wise regression with heterogeneous data. *J. Bus. Econom. Statist.* **41** 1157–1172. MR4650452 https://doi.org/10.1080/07350015.2022.2115498

[29] LI, X. and MENG, X.-L. (2021). A multi-resolution theory for approximating infinite-p-zero-n: Transitional inference, individualized predictions, and a world without bias-variance tradeoff. *J. Amer. Statist. Assoc.* **116** 353–367. MR4227699 https://doi.org/10.1080/01621459.2020.1844210

[30] LIN, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.* **28** 734–755. MR1792785 https://doi.org/10.1214/aos/1015951996

[31] MCCULLAGH, P. (2000). Invariance and factorial models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 209–256. With discussion and a reply by the author. MR1749537 https://doi.org/10.1111/1467-9868.00229

[32] MCCULLAGH, P. and NELDER, J. A. (2019). *Generalized Linear Models*. Routledge. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

[33] MURDOCH, W. J., SINGH, C., KUMBIER, K., ABBASI-ASL, R. and YU, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **116** 22071–22080. MR4030584 https://doi.org/10.1073/pnas.1900654116

[34] O'DONNELL, R. (2014). *Analysis of Boolean Functions*. Cambridge Univ. Press, New York. MR3443800 https://doi.org/10.1017/CBO9781139814782

[35] RAINEY, C. (2016). Compression and conditional effects: A product term is essential when using logistic regression to test for interaction. *Political Science Research and Methods* **4** 621–639.

[36] RAINEY, C. (2016). Dealing with separation in logistic regression models. *Polit. Anal.* **24** 339–355.

[37] RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343 https://doi.org/10.1214/09-AOS691

[38] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[39] RUDIN, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1** 206–215. https://doi.org/10.1038/s42256-019-0048-x

[40] RUSSETT, B. M. and ONEAL, J. R. (2001). *Triangulating Peace: Democracy, Interdependence, and International Organizations*. W. W. Norton, New York.

[41] SPECKMAN, P. L., LEE, J. and SUN, D. (2009). Existence of the MLE and propriety of posteriors for a general multinomial choice model. *Statist. Sinica* **19** 731–748. MR2514185

[42] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

[43] THANEI, G.-A., MEINSHAUSEN, N. and SHAH, R. D. (2018). The xyz algorithm for fast interaction search in high-dimensional data. *J. Mach. Learn. Res.* **19** Paper No. 37, 42. MR3862444

[44] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242

[45] VALIANT, L. G. (1984). A theory of the learnable. *Commun. ACM* **27** 1134–1142.

[46] VANDERWEELE, T. J. and ROBINS, J. M. (2008). Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika* **95** 49–61. MR2409714 https://doi.org/10.1093/biomet/asm090

[47] VANSTEELANDT, S. and DUKES, O. (2022). Assumption-lean inference for generalised linear model parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 657–685. MR4460549

[48] WAHBA, G., WANG, Y., GU, C., KLEIN, R. and KLEIN, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. *Ann. Statist.* **23** 1865–1895. MR1389856 https://doi.org/10.1214/aos/1034713638

[49] WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

[50] XIANG, S., ZHANG, W., LIU, S., HOADLEY, K. A., PEROU, C. M., ZHANG, K. and MARRON, J. S. (2023). Pairwise nonlinear dependence analysis of genomic data. *Ann. Appl. Stat.* **17** 2924–2943. MR4661682 https://doi.org/10.1214/23-aoas1745

[51] ZHANG, K. (2017). Spherical cap packing asymptotics and rank-extreme detection. *IEEE Trans. Inf. Theory* **63** 4572–4584. MR3666977 https://doi.org/10.1109/TIT.2017.2700202

[52] ZHANG, K. (2019). BET on independence. *J. Amer. Statist. Assoc.* **114** 1620–1637. MR4047288 https://doi.org/10.1080/01621459.2018.1537921

[53] ZHANG, K., ZHANG, W., ZHAO, Z. and ZHOU, W. (2021). BEAUTY powered BEAST. arXiv preprint arXiv:2103.00674.

[54] ZHANG, W., ZHAO, Z., BAIOCCHI, M., LI, Y. and ZHANG, K. (2021). SorBET: A Fast and Powerful Algorithm to Test Dependence of Variables. Submitted.

The Annals of Statistics
2025, Vol. 53, No. 3, 1095–1122
https://doi.org/10.1214/25-AOS2494
© Institute of Mathematical Statistics, 2025

SPARSITY MEETS CORRELATION IN GAUSSIAN SEQUENCE MODEL

BY SUBHODH KOTEKAL[a] AND CHAO GAO[b]

Department of Statistics, University of Chicago, [a]*skotekal@uchicago.edu,* [b]*chaogao@uchicago.edu*

We study estimation of an s-sparse signal in the p-dimensional Gaussian sequence model with equicorrelated observations and derive the minimax rate. A new phenomenon emerges from correlation, namely, the rate scales with respect to $p-2s$ and exhibits a phase transition at $p-2s \asymp \sqrt{p}$. Correlation is shown to be a blessing, provided it is sufficiently strong and the critical correlation level exhibits a delicate dependence on the sparsity level. Due to correlation, the minimax rate is driven by two subproblems: estimation of a linear functional (the average of the signal) and estimation of the signal's $(p-1)$-dimensional projection onto the orthogonal subspace. The high-dimensional projection is estimated via sparse regression, and the linear functional is cast as a robust location estimation problem. Existing robust estimators turn out to be suboptimal, and we show a kernel mode estimator with a widening bandwidth exploits the Gaussian character of the data to achieve the optimal estimation rate.

REFERENCES

[1] ANTONIADIS, A. (2007). Wavelet methods in statistics: Some recent developments and their applications. *Stat. Surv.* **1** 16–55. MR2520413 https://doi.org/10.1214/07-SS014

[2] ARIAS-CASTRO, E., QIAO, W. and ZHENG, L. (2022). Estimation of the global mode of a density: Minimaxity, adaptation, and computational complexity. *Electron. J. Stat.* **16** 2774–2795. MR4408293 https://doi.org/10.1214/21-ejs1972

[3] BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315

[4] BELLEC, P. C., LECUÉ, G. and TSYBAKOV, A. B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. MR3852663 https://doi.org/10.1214/17-AOS1670

[5] BERAN, J. (1992). Statistical methods for data with long-range dependence. *Statist. Sci.* **7** 404–416. https://doi.org/10.1214/ss/1177011122

[6] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[7] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

[8] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[9] CAI, T. T. and JIN, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* **38** 100–145. MR2589318 https://doi.org/10.1214/09-AOS696

[10] CAI, T. T. and LOW, M. G. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552–576. MR2060169 https://doi.org/10.1214/009053604000000094

[11] CAI, T. T. and LOW, M. G. (2006). Adaptation under probabilistic error for estimating linear functionals. *J. Multivariate Anal.* **97** 231–245. MR2208851 https://doi.org/10.1016/j.jmva.2005.03.001

[12] CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.* **39** 1012–1041. MR2816346 https://doi.org/10.1214/10-AOS849

[13] CARPENTIER, A., DELATTRE, S., ROQUAIN, E. and VERZELEN, N. (2021). Estimating minimum effect with outlier selection. *Ann. Statist.* **49** 272–294. MR4206678 https://doi.org/10.1214/20-AOS1956

MSC2020 subject classifications. 62C20, 62G05, 62G35.
Key words and phrases. High-dimensional statistics, minimax rate, correlation, mode estimation.

[14] CARPENTIER, A. and VERZELEN, N. (2019). Adaptive estimation of the sparsity in the Gaussian vector model. *Ann. Statist.* **47** 93–126. MR3909928 https://doi.org/10.1214/17-AOS1680
[15] CHEN, M., GAO, C. and REN, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electron. J. Stat.* **10** 3752–3774. MR3579675 https://doi.org/10.1214/16-EJS1216
[16] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 https://doi.org/10.1007/BF02868560
[17] COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432
[18] COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and VERZELEN, N. (2018). Optimal adaptive estimation of linear functionals under sparsity. *Ann. Statist.* **46** 3130–3150. MR3851767 https://doi.org/10.1214/17-AOS1653
[19] COLLIER, O. and DALALYAN, A. S. (2018). Estimating linear functionals of a sparse family of Poisson means. *Stat. Inference Stoch. Process.* **21** 331–344. MR3824971 https://doi.org/10.1007/s11203-018-9173-0
[20] COLLIER, O. and DALALYAN, A. S. (2019). Multidimensional linear functional estimation in sparse Gaussian models and robust estimation of the mean. *Electron. J. Stat.* **13** 2830–2864. MR3998929 https://doi.org/10.1214/19-EJS1590
[21] DALALYAN, A. and THOMPSON, P. (2019). Outlier-robust estimation of a sparse linear model using ℓ_1-penalized Huber's M-estimator. In *Advances in Neural Information Processing Systems* **32**.
[22] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. MR1105839 https://doi.org/10.1214/aos/1176348114
[23] EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882. MR0572631
[24] EFRON, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99** 96–104. MR2054289 https://doi.org/10.1198/016214504000000089
[25] EFRON, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Amer. Statist. Assoc.* **102** 93–103. MR2293302 https://doi.org/10.1198/016214506000001211
[26] EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1–22. MR2431866 https://doi.org/10.1214/07-STS236
[27] FOYGEL, R. and MACKEY, L. (2014). Corrupted sensing: Novel guarantees for separating structured signals. *IEEE Trans. Inf. Theory* **60** 1223–1247. MR3164972 https://doi.org/10.1109/TIT.2013.2293654
[28] GANNAZ, I. (2007). Robust estimation and wavelet thresholding in partially linear models. *Stat. Comput.* **17** 293–310. MR2409795 https://doi.org/10.1007/s11222-007-9019-x
[29] GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653 https://doi.org/10.3150/07-BEJ110
[30] HALL, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587 https://doi.org/10.1214/aos/1176348651
[31] HALL, P. and HART, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351. MR1084984 https://doi.org/10.1016/0304-4149(90)90100-7
[32] IBRAGIMOV, I. A. and KHASMINSKIĬ, R. Z. (1984). Nonparametric estimation of the value of a linear functional in Gaussian white noise. *Teor. Veroyatn. Primen.* **29** 19–32. MR0739497
[33] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8
[34] JIANG, H. (2017). Uniform convergence rates for kernel density estimation. In *Proceedings of the* 34*th International Conference on Machine Learning* 1694–1703.
[35] JOHNSTONE, I. M. Gaussian estimation: sequence and wavelet models. Available at https://imjohnstone.su.domains/GE_08_09_17.pdf.
[36] JOHNSTONE, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statist. Sinica* **9** 51–83. MR1678881
[37] JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B, Methodol.* **59** 319–351. MR1440585 https://doi.org/10.1111/1467-9868.00071
[38] KOTEKAL, S. and GAO, C. (2023). Minimax rates for sparse signal detection under correlation. *Inf. Inference* **12** Paper No. iaad044, 97. MR4669791 https://doi.org/10.1093/imaiai/iaad044
[39] KOTEKAL, S. and GAO, C. (2025). Supplement to "Sparsity meets correlation in Gaussian sequence model." https://doi.org/10.1214/25-AOS2494SUPP
[40] LEEK, J. T. and STOREY, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3** e161. https://doi.org/10.1371/journal.pgen.0030161.

[41] LEEK, J. T. and STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. USA* **105** 18718–18723. https://doi.org/10.1073/pnas.0808709105

[42] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006

[43] LIU, H. and GAO, C. (2019). Density estimation with contamination: Minimax rates and theory of adaptation. *Electron. J. Stat.* **13** 3613–3653. MR4013747 https://doi.org/10.1214/19-ejs1617

[44] LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

[45] MCCANN, L. and WELSCH, R. E. (2007). Robust variable selection using least angle regression and elemental set sampling. *Comput. Statist. Data Anal.* **52** 249–257. MR2409979 https://doi.org/10.1016/j.csda.2007.01.012

[46] MCCLEAN, A., BALAKRISHNAN, S., KENNEDY, E. H. and WASSERMAN, L. (2024). Double cross-fit doubly robust estimators: Beyond series regression. arXiv preprint. Available at arXiv:2403.15175.

[47] MCGRATH, S. and MUKHERJEE, R. (2005). Nuisance function tuning and sample splitting for optimal doubly robust estimation. arXiv preprint. Available at arXiv:2212.14857.

[48] MELNYK, I. and BANERJEE, A. (2016). Estimating structured vector autoregressive models. In *Proceedings of the* 33*rd International Conference on Machine Learning* 830–839.

[49] NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133 https://doi.org/10.1214/12-STS400

[50] NEWEY, W. K. and ROBINS, J. R. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. arXiv preprint. Available at arXiv:1801.09138.

[51] NGUYEN, N. H. and TRAN, T. D. (2013). Robust Lasso with missing and grossly corrupted observations. *IEEE Trans. Inf. Theory* **59** 2036–2058. MR3043781 https://doi.org/10.1109/TIT.2012.2232347

[52] OPSOMER, J., WANG, Y. and YANG, Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.* **16** 134–153. MR1861070 https://doi.org/10.1214/ss/1009213287

[53] PANINSKI, L. and YAJIMA, M. (2008). Undersmoothed kernel entropy estimators. *IEEE Trans. Inf. Theory* **54** 4384–4388. MR2451978 https://doi.org/10.1109/TIT.2008.928251

[54] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076. MR0143282 https://doi.org/10.1214/aoms/1177704472

[55] QIU, X., BROOKS, A. I., KLEBANOV, L. and YAKOVLEV, N. (2005). The effects of normalization on the correlation structure of microarray data. *BMC Bioinform.* **6** 1–11. https://doi.org/10.1186/1471-2105-6-120

[56] QIU, X., KLEBANOV, L. and YAKOVLEV, A. (2005). Correlation between gene expression levels and limitations of the empirical Bayes methodology for finding differentially expressed genes. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 34, 32. MR2183944 https://doi.org/10.2202/1544-6115.1157

[57] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837. MR0079873 https://doi.org/10.1214/aoms/1177728190

[58] SHU, H. and NAN, B. (2019). Estimation of large covariance and precision matrices from temporally dependent observations. *Ann. Statist.* **47** 1321–1350. MR3911114 https://doi.org/10.1214/18-AOS1716

[59] SUN, L. and STEPHENS, M. (2018). Solving the empirical Bayes normal means problem with correlated noise. arXiv preprint. Available at arXiv:1812.07488. https://doi.org/10.48550/arXiv.1812.07488

[60] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[61] VERSHYNIN, R. (2018). *High-Dimensional Probability*: *An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

[62] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

[63] WANG, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24** 466–484. MR1394972 https://doi.org/10.1214/aos/1032894449

[64] WU, Y. and YANG, P. (2020). Polynomial methods in statistical inference: Theory and practice. *Found. Trends Commun. Inf. Theory* **17** 402–586. https://doi.org/10.1561/0100000095

[65] ZHANG, P. and REN, Z. (2023). Adaptive minimax density estimation on $\mathbb{R}^d$ for Huber's contamination model. *Inf. Inference* **12** Paper No. iaad045, 25. MR4670354 https://doi.org/10.1093/imaiai/iaad045

The Annals of Statistics
2025, Vol. 53, No. 3, 1123–1149
https://doi.org/10.1214/25-AOS2495
© Institute of Mathematical Statistics, 2025

CONFORMAL INFERENCE FOR RANDOM OBJECTS

BY HANG ZHOU[a] AND HANS-GEORG MÜLLER[b]

Department of Statistics, University of California, Davis, [a]*hgzhou@ucdavis.edu,* [b]*hgmueller@ucdavis.edu*

We develop an inferential tool kit for analyzing object-valued responses, which correspond to data situated in general metric spaces, paired with Euclidean predictors within the conformal framework. To this end, we introduce conditional profile average transport costs, where we compare distance profiles that correspond to one-dimensional distributions of probability mass falling into balls of increasing radius through the optimal transport cost when moving from one distance profile to another. The average transport cost to transport a given distance profile to all others is crucial for statistical inference in metric spaces and underpins the proposed conditional profile scores. A key feature of the proposed approach is to utilize the distribution of conditional profile average transport costs as conformity score for general metric space-valued responses, which facilitates the construction of prediction sets by the split conformal algorithm. We derive the uniform convergence rate of the proposed conformity score estimators and establish asymptotic conditional validity for the prediction sets. The finite sample performance for synthetic data in various metric spaces demonstrates that the proposed conditional profile score outperforms existing methods in terms of both coverage level and size of the resulting prediction sets, even in the special case of scalar Euclidean responses. We also demonstrate the practical utility of conditional profile scores for network data from New York taxi trips and for compositional data reflecting energy sourcing of U.S. states.

REFERENCES

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

ANGELOPOULOS, A., CANDÈS, E. and TIBSHIRANI, R. J. (2024). Conformal PID control for time series prediction. In *Adv. Neural Inf. Process. Syst* 36.

BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). The limits of distribution-free conditional predictive inference. *Inf. Inference* **10** 455–482. MR4270755 https://doi.org/10.1093/imaiai/iaaa017

BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2023). Conformal prediction beyond exchangeability. *Ann. Statist.* **51** 816–845. MR4601003 https://doi.org/10.1214/23-aos2276

BATES, S., CANDÈS, E., LEI, L., ROMANO, Y. and SESIA, M. (2023). Testing for outliers with conformal p-values. *Ann. Statist.* **51** 149–178. MR4564852 https://doi.org/10.1214/22-aos2244

BHATTACHARJEE, S. and MÜLLER, H.-G. (2023). Single index Fréchet regression. *Ann. Statist.* **51** 1770–1798. MR4658576 https://doi.org/10.1214/23-aos2307

BIGOT, J., GOUET, R., KLEIN, T. and LÓPEZ, A. (2017). Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1–26. MR3606732 https://doi.org/10.1214/15-AIHP706

BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

CANDÈS, E., LEI, L. and REN, Z. (2023). Conformalized survival analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 24–45. MR4716146 https://doi.org/10.1093/jrsssb/qkac004

CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306. MR0981451 https://doi.org/10.1214/aos/1176347017

CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720–1747. MR2850218 https://doi.org/10.1214/11-AOS882

MSC2020 subject classifications. Primary 62R20, 62G20; secondary 62G99, 62J99.
Key words and phrases. Conditional distance profiles, conformity score, empirical process, non-Euclidean data, transport cost, uniform convergence.

CHEN, Y., LIN, Z. and MÜLLER, H.-G. (2023). Wasserstein regression. *J. Amer. Statist. Assoc.* **118** 869–882. MR4595462 https://doi.org/10.1080/01621459.2021.1956937

CHERNOZHUKOV, V., WÜTHRICH, K. and ZHU, Y. (2021). Distributional conformal prediction. *Proc. Natl. Acad. Sci. USA* **118** Paper No. e2107794118, 9. MR4389990 https://doi.org/10.1073/pnas.2107794118

CHOI, E. and HALL, P. (1998). On bias reduction in local linear smoothing. *Biometrika* **85** 333–345. MR1649117 https://doi.org/10.1093/biomet/85.2.333

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

DU, L., GUO, X., SUN, W. and ZOU, C. (2023). False discovery rate control under general dependence by symmetrized data aggregation. *J. Amer. Statist. Assoc.* **118** 607–621. MR4571145 https://doi.org/10.1080/01621459.2021.1945459

DUBEY, P., CHEN, Y. and MÜLLER, H.-G. (2024). Metric statistics: Exploration and inference for random objects with distance profiles. *Ann. Statist.* **52** 757–792. MR4744195 https://doi.org/10.1214/24-aos2368

DUBEY, P. and MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. MR4084166

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216. MR1212173 https://doi.org/10.1214/aos/1176349022

FAN, J. and GIJBELS, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036. MR1193323 https://doi.org/10.1214/aos/1176348900

FERRATY, F., PARK, J. and VIEU, P. (2011). Estimation of a functional single index model. In *Recent Advances in Functional Data Analysis and Related Topics. Contrib. Statist.* 111–116. Physica-Verlag/Springer, Heidelberg. MR2815569 https://doi.org/10.1007/978-3-7908-2736-1_17

FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1993). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press, Cambridge. MR1247695

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017 https://doi.org/10.1007/s11263-012-0591-y

GAO, F. and WELLNER, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Sci. China Ser. A* **52** 1525–1538. MR2520591 https://doi.org/10.1007/s11425-009-0102-y

GIBBS, I. and CANDÈS, E. (2021). Adaptive conformal inference under distribution shift. In *Adv. Neural Inf. Process. Syst* **2021** 1660–1672.

GIBBS, I. and CANDÈS, E. (2024). Conformal inference for online prediction with arbitrary distribution shifts. *J. Mach. Learn. Res.* **25** Paper No. [162], 36. MR4758133

HALL, P. (1989). On projection pursuit regression. *Ann. Statist.* **17** 573–588. MR0994251 https://doi.org/10.1214/aos/1176347126

HALL, P. and MARRON, J. S. (1997). On the role of the shrinkage parameter in local linear smoothing. *Probab. Theory Related Fields* **108** 495–516. MR1465639 https://doi.org/10.1007/s004400050118

HALL, P., WOLFF, R. C. L. and YAO, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94** 154–163. MR1689221 https://doi.org/10.2307/2669691

HU, X. and LEI, J. (2024). A two-sample conditional distribution test using conformal prediction and weighted rank sum. *J. Amer. Statist. Assoc.* **119** 1136–1154. MR4765976 https://doi.org/10.1080/01621459.2023.2177165

ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981 https://doi.org/10.1016/0304-4076(93)90114-K

IZBICKI, R., SHIMIZU, G. and STERN, R. B. (2022). `CD-split` and `HPD-split`: Efficient conformal regions in high dimensions. *J. Mach. Learn. Res.* **23** Paper No. [87], 32. MR4576672

JIANG, C.-R. and WANG, J.-L. (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39** 362–388. MR2797850 https://doi.org/10.1214/10-AOS845

KUCHIBHOTLA, A. K. and PATRA, R. K. (2020). Efficient estimation in single index models through smoothing splines. *Bernoulli* **26** 1587–1618. MR4058379 https://doi.org/10.3150/19-BEJ1183

LEI, J., G'SELL, M., RINALDO, A., TIBSHIRANI, R. J. and WASSERMAN, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342 https://doi.org/10.1080/01621459.2017.1307116

LEI, J., ROBINS, J. and WASSERMAN, L. (2013). Distribution-free prediction sets. *J. Amer. Statist. Assoc.* **108** 278–287. MR3174619 https://doi.org/10.1080/01621459.2012.751873

LEI, J. and WASSERMAN, L. (2014). Distribution-free prediction bands for non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 71–96. MR3153934 https://doi.org/10.1111/rssb.12021

LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. MR4352539 https://doi.org/10.1214/21-aos2095

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584 https://doi.org/10.1198/jasa.2009.tm08647

PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. MR3449766 https://doi.org/10.1214/15-AOS1363

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

PETERSEN, A., ZHANG, C. and KOKOSZKA, P. (2022). Modeling probability density functions as data objects. *Econom. Stat.* **21** 159–178. MR4366852 https://doi.org/10.1016/j.ecosta.2021.04.004

ROMANO, Y., PATTERSON, E. and CANDÈS, E. (2019). Conformalized quantile regression. In *Adv. Neural Inf. Process. Syst* 32.

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

SESIA, M. and CANDÈS, E. J. (2020). A comparison of some conformal quantile regression methods. *Stat* **9** e261, 8. MR4104217 https://doi.org/10.1002/sta4.261

VILLANI, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin.

VOVK, V. (2021). Conformal testing in a binary model situation. *Proc. Mach. Learn. Res.* **152** 131–150.

VOVK, V. (2013). Conditional validity of inductive conformal predictors. *Mach. Learn.* **92** 349–376. MR3080332 https://doi.org/10.1007/s10994-013-5355-6

VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

VOVK, V., NOURETDINOV, I. and GAMMERMAN, A. (2009). On-line predictive linear regression. *Ann. Statist.* **37** 1566–1590. MR2509084 https://doi.org/10.1214/08-AOS622

VOVK, V., PETEJ, I., NOURETDINOV, I., AHLBERG, E., CARLSSON, L. and GAMMERMAN, A. (2021). Retrain or not retrain: Conformal test martingales for change-point detection. In *Conformal and Probabilistic Prediction and Applications* 191–210.

WANG, X., ZHU, J., PAN, W., ZHU, J. and ZHANG, H. (2024). Nonparametric statistical inference via metric distribution function in metric spaces. *J. Amer. Statist. Assoc.* **119** 2772–2784. MR4833914 https://doi.org/10.1080/01621459.2023.2277417

WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. MR2543689 https://doi.org/10.1214/08-AOS646

YANG, Z., CANDÈS, E. and LEI, L. (2024). Bellman conformal inference: Calibrating prediction intervals for time series. Preprint. Available at arXiv:2402.05203.

YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x

ZHOU, J. and HE, X. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *Ann. Statist.* **36** 1649–1668. MR2435451 https://doi.org/10.1214/07-AOS529

ZHOU, H. and MÜLLER, H.-G (2025). Supplement to "Conformal inference for random objects." https://doi.org/10.1214/25-AOS2495SUPPA, https://doi.org/10.1214/25-AOS2495SUPPB

ZHU, C. and MÜLLER, H.-G. (2023). Autoregressive optimal transport models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1012–1033. MR4726980 https://doi.org/10.1093/jrsssb/qkad051

ZHU, L.-P. and ZHU, L.-X. (2009). On distribution-weighted partial least squares with diverging number of highly correlated predictors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 525–548. MR2649607 https://doi.org/10.1111/j.1467-9868.2008.00697.x

ZOU, C., WANG, G. and LI, R. (2020). Consistent selection of the number of change-points via sample-splitting. *Ann. Statist.* **48** 413–439. MR4065168 https://doi.org/10.1214/19-AOS1814

The Annals of Statistics
2025, Vol. 53, No. 3, 1150–1175
https://doi.org/10.1214/25-AOS2496
© Institute of Mathematical Statistics, 2025

STATISTICAL ALGORITHMS FOR LOW-FREQUENCY DIFFUSION DATA: A PDE APPROACH

BY MATTEO GIORDANO[1,a] AND SVEN WANG[2,b]

[1]*ESOMAS Department, University of Turin,* [a]*matteo.giordano@unito.it*
[2]*Department of Mathematics, Humboldt University Berlin,* [b]*sven.wang@hu-berlin.de*

We consider the problem of making nonparametric inference in a class of multi-dimensional diffusions in divergence form, from low-frequency data. Statistical analysis in this setting is notoriously challenging due to the intractability of the likelihood and its gradient, and computational methods have thus far largely resorted to expensive simulation-based techniques. In this article, we propose a new computational approach which is motivated by PDE theory and is built around the characterisation of the transition densities as solutions of the associated heat (Fokker-Planck) equation. Employing optimal regularity results from the theory of parabolic PDEs, we prove a novel characterisation for the gradient of the likelihood. Using these developments, for the nonlinear inverse problem of recovering the diffusivity, we then show that the numerical evaluation of the likelihood and its gradient can be reduced to standard elliptic eigenvalue problems, solvable by powerful finite element methods. This enables the efficient implementation of a large class of popular statistical algorithms, including (i) preconditioned Crank-Nicolson and Langevin-type methods for posterior sampling, and (ii) gradient-based descent optimisation schemes to compute maximum likelihood and maximum-a-posteriori estimates. We showcase the effectiveness of these methods via extensive simulation studies in a nonparametric Bayesian model with Gaussian process priors, in which both the proposed optimisation and sampling schemes provide good numerical recovery.

REFERENCES

[1] ABRAHAM, K. and NICKL, R. (2019). On statistical Calderón problems. *Math. Stat. Learn.* **2** 165–216. MR4130599

[2] AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** 223–262. MR1926260 https://doi.org/10.1111/1468-0262.00274

[3] AÏT-SAHALIA, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36** 906–937. MR2396819 https://doi.org/10.1214/009053607000000622

[4] ALBERTI, G. S., BARNES, D., JAMBHALE, A. and NICKL, R. (2024). On low frequency inference for diffusions without the hot spots conjecture. arXiv preprint. Available at arXiv:2410.19393.

[5] ALLEN, E. (2007). *Modeling with Itô Stochastic Differential Equations*. *Mathematical Modelling: Theory and Applications* **22**. Springer, Dordrecht. MR2292765

[6] ALTMEYER, R. (2022). Polynomial time guarantees for sampling based posterior inference in high-dimensional generalised linear models. arXiv preprint. Available at arXiv:2208.13296.

[7] ARONSON, D. G. (1967). Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73** 890–896. MR0217444 https://doi.org/10.1090/S0002-9904-1967-11830-5

[8] BABUŠKA, I. and OSBORN, J. (1991). Eigenvalue problems. In *Handbook of Numerical Analysis, Vol. II*. *Handb. Numer. Anal., II* 641–787. North-Holland, Amsterdam. MR1115240

[9] BABUŠKA, I. and OSBORN, J. E. (1989). Finite element-Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems. *Math. Comp.* **52** 275–297. MR0962210 https://doi.org/10.2307/2008468

MSC2020 subject classifications. Primary 62M15; secondary 62F15, 62G05.
Key words and phrases. Low-frequency data, inference for SDEs, nonparametric Bayesian inference, Markov chain Monte Carlo, gradient-based optimisation, numerical methods for PDEs, intractable likelihood.

[10] BAEHR, H. D. and STEPHAN, K. (1998). Heat conduction and mass diffusion. In *Heat and Mass Transfer* 105–250. Springer, Berlin.

[11] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **348**. Springer, Cham. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

[12] BANDEIRA, A. S., MAILLARD, A., NICKL, R. and WANG, S. (2023). On free energy barriers in Gaussian priors and failure of cold start MCMC for high-dimensional unimodal distributions. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **381** 20220150. MR4590496 https://doi.org/10.1098/rsta.2022.0150

[13] BESKOS, A., GIROLAMI, M., LAN, S., FARRELL, P. E. and STUART, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* **335** 327–351. MR3612501 https://doi.org/10.1016/j.jcp.2016.12.041

[14] BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *Ann. Statist.* **37** 223–245. MR2488350 https://doi.org/10.1214/07-AOS550

[15] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331 https://doi.org/10.1111/j.1467-9868.2006.00552.x

[16] BLADT, M., FINCH, S. and SØRENSEN, M. (2016). Simulation of multivariate diffusion bridges. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 343–369. MR3454200 https://doi.org/10.1111/rssb.12118

[17] BOFFI, D. (2010). Finite element approximation of eigenvalue problems. *Acta Numer.* **19** 1–120. MR2652780 https://doi.org/10.1017/S0962492910000012

[18] BOHR, J. and NICKL, R. (2024). On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. *Ann. Inst. Henri Poincaré Probab. Stat.* **60** 2619–2667. MR4828853 https://doi.org/10.1214/23-aihp1397

[19] BRAMBLE, J. H. and OSBORN, J. E. (1973). Rate of convergence estimates for nonselfadjoint eigenvalue approximations. *Math. Comp.* **27** 525–549. MR0366029 https://doi.org/10.2307/2005658

[20] BRIANE, V., VIMOND, M. and KERVRANN, C. (2020). An overview of diffusion models for intracellular dynamics analysis. *Brief. Bioinform.* **21** 1136–1150. https://doi.org/10.1093/bib/bbz052

[21] CALDERÓN, A.-P. (1980). On an inverse boundary value problem. In *Seminar on Numerical Analysis and Its Applications to Continuum Physics* (*Rio de Janeiro*, 1980) 65–73. Soc. Brasil. Mat., Rio de Janeiro. MR0590275

[22] CAO, H., TAN, C., GAO, Z., XU, Y., CHEN, G., HENG, P.-A. and LI, S. Z. (2024). A survey on generative diffusion models. *IEEE Trans. Knowl. Data Eng.*.

[23] CHATELIN, F. (1973). Convergence of approximation methods to compute eigenelements of linear operations. *SIAM J. Numer. Anal.* **10** 939–948. MR0349004 https://doi.org/10.1137/0710080

[24] COMTE, F., GENON-CATALOT, V. and ROZENHOLC, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** 514–543. MR2331262 https://doi.org/10.3150/07-BEJ5173

[25] COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540 https://doi.org/10.1214/13-STS421

[26] CUI, T., LAW, K. J. H. and MARZOUK, Y. M. (2016). Dimension-independent likelihood-informed MCMC. *J. Comput. Phys.* **304** 109–137. MR3422405 https://doi.org/10.1016/j.jcp.2015.10.008

[27] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451 https://doi.org/10.1007/s00440-006-0502-7

[28] DARCY, M., HAMZI, B., LIVIERI, G., OWHADI, H. and TAVALLALI, P. (2023). One-shot learning of stochastic differential equations with data adapted kernels. *Phys. D* **444** 133583. MR4517044 https://doi.org/10.1016/j.physd.2022.133583

[29] DAVIES, E. B. (1995). *Spectral Theory and Differential Operators*. *Cambridge Studies in Advanced Mathematics* **42**. Cambridge Univ. Press, Cambridge. MR1349825 https://doi.org/10.1017/CBO9780511623721

[30] DELYON, B. and HU, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Process. Appl.* **116** 1660–1675. MR2269221 https://doi.org/10.1016/j.spa.2006.04.004

[31] DESCLOUX, J., NASSIF, N. and RAPPAZ, J. (1978). On spectral approximation. I. The problem of convergence. *RAIRO. Anal. Numér.* **12** 97–112. MR0483400 https://doi.org/10.1051/m2an/1978120200971

[32] DEXHEIMER, N., STRAUCH, C. and TROTTNER, L. (2022). Adaptive invariant density estimation for continuous-time mixing Markov processes under sup-norm risk. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 2029–2064. MR4492970 https://doi.org/10.1214/21-aihp1235

[33] DURHAM, G. B. and GALLANT, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** 297–338. MR1939904 https://doi.org/10.1198/073500102288618397

[34] ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. MR1839375 https://doi.org/10.1111/1468-0262.00226

[35] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems. Mathematics and Its Applications* **375**. Kluwer Academic, Dordrecht. MR1408680

[36] ERAKER, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19** 177–191. MR1939708 https://doi.org/10.1198/073500101316970403

[37] EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR2597943 https://doi.org/10.1090/gsm/019

[38] FOULON, S., TRUONG, T., LEUTENEGGER, A.-L. and PERDRY, H. (2024). Parameter estimation of hidden Markov models: Comparison of EM and quasi-Newton methods with a new hybrid algorithm. arXiv preprint. Available at arXiv:2409.02477.

[39] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[40] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

[41] GIORDANO, M. and NICKL, R. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Probl.* **36** 085001. MR4151406 https://doi.org/10.1088/1361-6420/ab7d2a

[42] GIORDANO, M. and RAY, K. (2022). Nonparametric Bayesian inference for reversible multidimensional diffusions. *Ann. Statist.* **50** 2872–2898. MR4500628 https://doi.org/10.1214/22-aos2213

[43] GIORDANO, M. and WANG, S. (2025). Supplement to "Statistical algorithms for low-frequency diffusion data: a PDE approach." https://doi.org/10.1214/25-AOS2496SUPP

[44] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509 https://doi.org/10.1214/009053604000000797

[45] GOLIGHTLY, A. and WILKINSON, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52** 1674–1693. MR2422763 https://doi.org/10.1016/j.csda.2007.05.019

[46] HAIRER, M., STUART, A. M. and VOLLMER, S. J. (2014). Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. MR3262508 https://doi.org/10.1214/13-AAP982

[47] HECKERT, A., DAHAL, L., TJIAN, R. and DARZACQ, X. (2022). Recovering mixtures of fast-diffusing states from short single-particle trajectories. *eLife* **11**. https://doi.org/10.7554/eLife.70169

[48] HELTBERG, M. L., MINÉ-HATTAB, J., TADDEI, A., WALCZAK, A. M. and MORA, T. (2021). Physical observables to determine the nature of membrane-less cellular sub-compartments. *eLife* **10**. https://doi.org/10.7554/eLife.69181

[49] HOFFMANN, M. and RAY, K. (2024). Bayesian estimation in a multidimensional diffusion model with high frequency data. *Probab. Theory Related Fields*.

[50] KESSLER, M. and SØRENSEN, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314. MR1681700 https://doi.org/10.2307/3318437

[51] KNYAZEV, A. V. and OSBORN, J. E. (2006). New a priori FEM error estimates for eigenvalues. *SIAM J. Numer. Anal.* **43** 2647–2667. MR2206452 https://doi.org/10.1137/040613044

[52] KOHN, R. and VOGELIUS, M. (1984). Determining conductivity by boundary measurements. *Comm. Pure Appl. Math.* **37** 289–298. MR0739921 https://doi.org/10.1002/cpa.3160370302

[53] LEWIS, M. A., MAINI, P. K. and PETROVSKII, S. V. (2013). Dispersal, individual movement and spatial ecology. *Lecture Notes in Math.* **2071**.

[54] LIN, M., CHEN, R. and MYKLAND, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105** 820–838. MR2724864 https://doi.org/10.1198/jasa.2010.tm09057

[55] LOU, A. and ERMON, S. (2023). Reflected diffusion models. In *International Conference on Machine Learning* 22675–22701. PMLR.

[56] LUNARDI, A. (1995). *Analytic Semigroups and Optimal Regularity in Parabolic Problems. Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. MR3012216

[57] MAJDA, A. J. and HARLIM, J. (2012). *Filtering Complex Turbulent Systems*. Cambridge Univ. Press, Cambridge. MR2934167 https://doi.org/10.1017/CBO9781139061308

[58] MBALAWATA, I. S., SÄRKKÄ, S. and HAARIO, H. (2013). Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering. *Comput. Statist.* **28** 1195–1223. MR3062727 https://doi.org/10.1007/s00180-012-0352-y

[59] NACHMAN, A. I. (1988). Reconstructions from boundary measurements. *Ann. of Math.* (2) **128** 531–576. MR0970610 https://doi.org/10.2307/1971435

[60] NICKL, R. (2023). *Bayesian Non-linear Statistical Inverse Problems*. *Zurich Lectures in Advanced Mathematics*. EMS Press.

[61] NICKL, R. (2024). Consistent inference for diffusions from low frequency measurements. *Ann. Statist.* **52** 519–549. MR4744186 https://doi.org/10.1214/24-aos2357

[62] NICKL, R. and RAY, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.* **48** 1383–1408. MR4124327 https://doi.org/10.1214/19-AOS1851

[63] NICKL, R. and SÖHL, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693. MR3670192 https://doi.org/10.1214/16-AOS1504

[64] NICKL, R., VAN DE GEER, S. and WANG, S. (2020). Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J. Uncertain. Quantificat.* **8** 374–413. MR4074017 https://doi.org/10.1137/18M1236137

[65] NICKL, R. and WANG, S. (2024). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *J. Eur. Math. Soc.* (*JEMS*) **26** 1031–1112. MR4721029 https://doi.org/10.4171/jems/1304

[66] NORRIS, J. R. (1997). Long-time behaviour of heat flow: Global estimates and exact asymptotics. *Arch. Ration. Mech. Anal.* **140** 161–195. MR1482931 https://doi.org/10.1007/s002050050063

[67] OKO, K., AKIYAMA, S. and SUZUKI, T. (2023). Diffusion models are minimax optimal distribution estimators. In *International Conference on Machine Learning* 26517–26582. PMLR.

[68] PAPASPILIOPOULOS, O., ROBERTS, G. O. and STRAMER, O. (2013). Data augmentation for diffusions. *J. Comput. Graph. Statist.* **22** 665–688. MR3173736 https://doi.org/10.1080/10618600.2013.783484

[69] PEDERSEN, A. R. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1** 257–279. MR1363541 https://doi.org/10.2307/3318480

[70] PELETIER, M. A., SAVARÉ, G. and VENERONI, M. (2012). Chemical reactions as Γ-limit of diffusion [revised reprint of MR2679596]. *SIAM Rev.* **54** 327–352. MR2916310 https://doi.org/10.1137/110858781

[71] POKERN, Y., STUART, A. M. and VAN ZANTEN, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.* **123** 603–628. MR3003365 https://doi.org/10.1016/j.spa.2012.08.010

[72] RAMESH, A., DHARIWAL, P., NICHOL, A., CHU, C. and CHEN, M. (2022). Hierarchical text-conditional image generation with CLIP latents. arXiv e-prints arXiv–2204.

[73] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[74] RICHTER, G. R. (1981). An inverse problem for the steady state diffusion equation. *SIAM J. Appl. Math.* **41** 210–221. MR0628945 https://doi.org/10.1137/0141016

[75] ROBERTS, G. O. and STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88** 603–621. MR1859397 https://doi.org/10.1093/biomet/88.3.603

[76] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

[77] ROMBACH, R., BLATTMANN, A., LORENZ, D., ESSER, P. and OMMER, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10684–10695.

[78] SALOFF-COSTE, L. (2010). The heat kernel and its estimates. In *Probabilistic Approach to Geometry*. *Adv. Stud. Pure Math.* **57** 405–436. Math. Soc. Japan, Tokyo. MR2648271 https://doi.org/10.2969/aspm/05710405

[79] SÄRKKÄ, S. and SOLIN, A. (2019). *Applied Stochastic Differential Equations*. *Institute of Mathematical Statistics Textbooks* **10**. Cambridge Univ. Press, Cambridge. MR3931353

[80] SCHAUER, M., VAN DER MEULEN, F. and VAN ZANTEN, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23** 2917–2950. MR3648050 https://doi.org/10.3150/16-BEJ833

[81] SHREVE, S. E. (2004). *Stochastic Calculus for Finance*. *II*. *Springer Finance*. Springer, New York. Continuous-time models. MR2057928

[82] SINGER, H. (2002). Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended Kalman filter and Itô-Taylor expansion. *J. Comput. Graph. Statist.* **11** 972–995. MR1944270 https://doi.org/10.1198/106186002321018894
[83] SONG, Y., SOHL-DICKSTEIN, J., KINGMA, D. P., KUMAR, A., ERMON, S. and POOLE, B. (2021). Score-based generative modeling through stochastic differential equations. In 9*th International Conference on Learning Representations*, *ICLR* 2021.
[84] STRAUCH, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46** 3451–3480. MR3852658 https://doi.org/10.1214/17-AOS1664
[85] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785 https://doi.org/10.1017/S0962492910000061
[86] TANAKA, H. (1979). Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.* **9** 163–177. MR0529332
[87] TAYLOR, M. E. (2011). *Partial Differential Equations I. Basic Theory*, 2nd ed. *Applied Mathematical Sciences* **115**. Springer, New York. MR2744150 https://doi.org/10.1007/978-1-4419-7055-8
[88] TIERNEY, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401 https://doi.org/10.1214/aoap/1027961031
[89] UHLMANN, G. (2009). Electrical impedance tomography and Calderón's problem. *Inverse Probl.* **25** 123011. MR3460047 https://doi.org/10.1088/0266-5611/25/12/123011
[90] VAINIKKO, G. M. (1964). Asymptotic error bounds for projection methods in the eigenvalue problem. *Zh. Vychisl. Mat. Mat. Fiz.* **4** 405–425. MR0176340
[91] VAN WAAIJ, J. and VAN ZANTEN, H. (2016). Gaussian process methods for one-dimensional diffusions: Optimal rates and adaptation. *Electron. J. Stat.* **10** 628–645. MR3471991 https://doi.org/10.1214/16-EJS1117
[92] VAN DER MEULEN, F. and SCHAUER, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11** 2358–2396. MR3656495 https://doi.org/10.1214/17-EJS1290
[93] VAN DER MEULEN, F. H., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* **12** 863–888. MR2265666 https://doi.org/10.3150/bj/1161614950
[94] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256
[95] VARADHAN, S. R. S. (1967). On the behavior of the fundamental solution of the heat equation with variable coefficients. *Comm. Pure Appl. Math.* **20** 431–455. MR0208191 https://doi.org/10.1002/cpa.3160200210
[96] WANG, S. (2019). The nonparametric LAN expansion for discretely observed diffusions. *Electron. J. Stat.* **13** 1329–1358.

The Annals of Statistics
2025, Vol. 53, No. 3, 1176–1202
https://doi.org/10.1214/25-AOS2497
© Institute of Mathematical Statistics, 2025

MINIMAX RATE FOR MULTIVARIATE DATA UNDER COMPONENTWISE LOCAL DIFFERENTIAL PRIVACY CONSTRAINTS

BY CHIARA AMORINO[1,a] AND ARNAUD GLOTER[2,b]

[1]*Department of Mathematics, Université du Luxembourg,* [a]*chiara.amorino@upf.edu*
[2]*Laboratoire de Mathématiques et Modélisation d'Evry, CNRS, Univ Evry, Université Paris-Saclay,* [b]*arnaud.gloter@univ-evry.fr*

Our research analyses the balance between maintaining privacy and preserving statistical accuracy when dealing with multivariate data that is subject to *componentwise local differential privacy* (CLDP). With CLDP, each component of the private data is made public through a separate privacy channel. This allows for varying levels of privacy protection for different components or for the privatization of each component by different entities, each with their own distinct privacy policies. It also covers the practical situations where it is impossible to privatize jointly all the components of the raw data. We develop general techniques for establishing minimax bounds that shed light on the statistical cost of privacy in this context, as a function of the privacy levels $\alpha_1, \dots, \alpha_d$ of the d components.

We demonstrate the versatility and efficiency of these techniques by presenting various statistical applications. Specifically, we examine nonparametric density and joint moments estimation under CLDP, providing upper and lower bounds that match up to constant factors, as well as an associated data-driven adaptive procedure. Additionally, we conduct a detailed analysis of the effective privacy level, exploring how information about a private characteristic of an individual may be inferred from the publicly visible characteristics of the same individual.

REFERENCES

[1] ABADI, M., CHU, A., GOODFELLOW, I., MCMAHAN, H. B., MIRONOV, I., TALWAR, K. and ZHANG, L. (2016). Deep learning with differential privacy. In *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security* 308–318.

[2] ACHARYA, J., SUN, Z. and ZHANG, H. (2021). Robust testing and estimation under manipulation attacks. In *International Conference on Machine Learning*. pp. 43–53. PMLR.

[3] AMORINO, C. and GLOTER, A. (2025). Supplement to "Minimax rate for multivariate data under componentwise local differential privacy constraints." https://doi.org/10.1214/25-AOS2497SUPP

[4] APPLE DIFFERENTIAL PRIVACY TEAM (2017). Learning with privacy at scale. Available at https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

[5] BERRETT, T. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *Adv. Neural Inf. Process. Syst.* **33** 3164–3173.

[6] BERRETT, T. and YU, Y. (2021). Locally private online change point detection. *Adv. Neural Inf. Process. Syst.* **34** 3425–3437.

[7] BERRETT, T. B., GYÖRFI, L. and WALK, H. (2021). Strongly universally consistent nonparametric regression and classification with privatised data. *Electron. J. Stat.* **15** 2430–2453. MR4255329 https://doi.org/10.1214/21-ejs1845

[8] BIE, A., KAMATH, G. and SINGHAL, V. (2022). Private estimation with public data. *Adv. Neural Inf. Process. Syst.* **35** 18653–18666.

[9] BOSQ, D. and BLANKE, D. (2007). *Inference and Prediction in Large Dimensions*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2364006 https://doi.org/10.1002/9780470724033

MSC2020 subject classifications. Primary 62C20, 62G07; secondary 62H05, 68P27.
Key words and phrases. Local differential privacy, minimax optimality, rate of convergence, non parametric estimation, density estimation.

[10] BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** 2524–2535. MR1425965 https://doi.org/10.1214/aos/1032181166

[11] BUTUCEA, C., DUBOIS, A., KROLL, M. and SAUMARD, A. (2020). Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli* **26** 1727–1764. MR4091090 https://doi.org/10.3150/19-BEJ1165

[12] BUTUCEA, C., ROHDE, A. and STEINBERGER, L. (2023). Interactive versus noninteractive locally differentially private estimation: Two elbows for the quadratic functional. *Ann. Statist.* **51** 464–486. MR4600989 https://doi.org/10.1214/22-aos2254

[13] DING, B., KULKARNI, J. and YEKHANIN, S. (2017). Collecting telemetry data privately. *Adv. Neural Inf. Process. Syst.* **30**.

[14] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114

[15] DUCHI, J. and ROGERS, R. (2019). Lower bounds for locally private estimation via communication complexity. In *Conference on Learning Theory*. pp. 1161–1191. PMLR.

[16] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy and statistical minimax rates. In 2013 *IEEE* 54*th Annual Symposium on Foundations of Computer Science—FOCS* 2013 429–438. IEEE Comput. Soc., Los Alamitos, CA. MR3246246 https://doi.org/10.1109/FOCS.2013.53

[17] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy, data processing inequalities, and statistical minimax rates. arXiv preprint. Available at arXiv:1302.3203.

[18] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

[19] DUCHI, J. C. and RUAN, F. (2024). The right complexity measure in locally private estimation: It is not the Fisher information. *Ann. Statist.* **52** 1–51. MR4718406 https://doi.org/10.1214/22-aos2227

[20] DUCHI, J. C., WAINWRIGHT, M. J. and JORDAN, M. I. (2013). Local privacy and minimax bounds: Sharp rates for probability estimation. *Adv. Neural Inf. Process. Syst.* **26**.

[21] DWORK, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*. *Lecture Notes in Computer Science* **4978** 1–19. Springer, Berlin. MR2472670 https://doi.org/10.1007/978-3-540-79228-4_1

[22] DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

[23] DWORK, C. and NISSIM, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology—CRYPTO* 2004: 24*th Annual International Cryptology Conference*, *Santa Barbara*, *California*, *USA*, *August* 15-19, 2004. *Proceedings*. *Lecture Notes in Computer Science* **3152** 528–544. Springer, Berlin. MR2147523 https://doi.org/10.1007/978-3-540-28628-8_32

[24] ERLINGSSON, U., PIHUR, V. and KOROLOVA, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the* 2014 *ACM SIGSAC Conference on Computer and Communications Security* 1054–1067.

[25] EVFIMIEVSKI, A., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 211–222.

[26] FORTI, M. (2021). The deployment of artificial intelligence tools in the health sector: Privacy concerns and regulatory answers within the GDPR. *Eur. J. Legal Stud.* **13** 29.

[27] GU, J. and CHEN, S. X. (2024). Statistical inference for decentralized federated learning. *Ann. Statist.* **52** 2931–2955. MR4842832 https://doi.org/10.1214/24-aos2452

[28] GYÖRFI, L. and KROLL, M. (2025). On rate optimal private regression under local differential privacy. *Statist. Sinica* **35** 613–627. MR4863657

[29] JOSEPH, M., MAO, J., NEEL, S. and ROTH, A. (2019). The role of interactivity in local differential privacy. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* (*FOCS*) 94–105. IEEE Comput. Soc., Los Alamitos, CA. MR4228162 https://doi.org/10.1109/FOCS.2019.00015

[30] KAIROUZ, P., OH, S. and VISWANATH, P. (2014). Extremal mechanisms for local differential privacy. In *Advances in Neural Information Processing Systems* **27**.

[31] KROLL, M. (2024). Adaptive spectral density estimation by model selection under local differential privacy. arXiv preprint. Available at arXiv:2010.04218.

[32] LAM-WEIL, J., LAURENT, B. and LOUBES, J.-M. (2022). Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli* **28** 579–600. MR4337717 https://doi.org/10.3150/21-bej1358

[33] LI, M., BERRETT, T. and YU, Y. (2022). Network change point localisation under local differential privacy. *Adv. Neural Inf. Process. Syst.* **35** 15013–15026.

[34] LI, M., BERRETT, T. B. and YU, Y. (2023). On robustness and local differential privacy. *Ann. Statist.* **51** 717–737. MR4600999 https://doi.org/10.1214/23-aos2267

[35] POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

[36] ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** 2646–2670. MR4152116 https://doi.org/10.1214/19-AOS1901

[37] SART, M. (2023). Density estimation under local differential privacy and Hellinger loss. *Bernoulli* **29** 2318–2341. MR4580918 https://doi.org/10.3150/22-bej1543

[38] TANG, J., KOROLOVA, A., BAI, X., WANG, X. and WANG, X. (2017). Privacy loss in apple's implementation of differential privacy on macos 10.12. arXiv preprint. Available at arXiv:1709.02753.

[39] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, Berlin. https://doi.org/10.1007/b13794

[40] WANG, D. and XU, J. (2021). On sparse linear regression in the local differential privacy model. *IEEE Trans. Inf. Theory* **67** 1182–1200. MR4232008 https://doi.org/10.1109/TIT.2020.3040406

[41] WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** 63–69.

[42] WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

The Annals of Statistics
2025, Vol. 53, No. 3, 1203–1229
https://doi.org/10.1214/25-AOS2500
© Institute of Mathematical Statistics, 2025

STRONG APPROXIMATIONS FOR EMPIRICAL PROCESSES INDEXED BY LIPSCHITZ FUNCTIONS

BY MATIAS D. CATTANEO[a] AND RUIQI RAE YU[b]

Department of Operations Research and Financial Engineering, Princeton University, [a]*cattaneo@princeton.edu,* [b]*rae.yu@princeton.edu*

This paper presents new uniform Gaussian strong approximations for empirical processes indexed by classes of functions based on d-variate random vectors ($d \geq 1$). First, a uniform Gaussian strong approximation is established for general empirical processes indexed by possibly Lipschitz functions, improving on previous results in the literature. In the setting considered by Rio (*Probab. Theory Related Fields* **98** (1994) 21–45), and if the function class is Lipschitzian, our result improves the approximation rate $n^{-1/(2d)}$ to $n^{-1/\max\{d,2\}}$, up to a polylog(n) term, where n denotes the sample size. Remarkably, we establish a valid uniform Gaussian strong approximation at the rate $n^{-1/2}\log n$ for $d = 2$, which was previously known to be valid only for univariate ($d = 1$) empirical processes via the celebrated Hungarian construction (Komlós, Major and Tusnády, *Z. Wahrsch. Verw. Gebiete* **32** (1975) 111–131). Second, a uniform Gaussian strong approximation is established for multiplicative separable empirical processes indexed by possibly Lipschitz functions, which addresses some outstanding problems in the literature (Chernozhukov, Chetverikov and Kato, *Ann. Statist.* **42** (2014) 1564–1597, Section 3). Finally, two other uniform Gaussian strong approximation results are presented when the function class is a sequence of Haar basis based on quasi-uniform partitions. Applications to nonparametric density and regression estimation are discussed.

REFERENCES

[1] AMBROSIO, L., FUSCO, N. and PALLARA, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. *Oxford Mathematical Monographs*. Oxford University Press, New York. MR1857292

[2] BECK, J. (1985). Lower bounds on the approximation of the multivariate empirical process. *Z. Wahrsch. Verw. Gebiete* **70** 289–306. MR0799151 https://doi.org/10.1007/BF02451433

[3] BERTHET, P. and MASON, D. M. (2006). Revisiting two strong approximation results of Dudley and Philipp. In *High Dimensional Probability*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 155–172. IMS, Beachwood, OH. MR2387767 https://doi.org/10.1214/074921706000000824

[4] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

[5] BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256. MR0972783

[6] BROWN, L. D., CAI, T. T. and ZHOU, H. H. (2010). Nonparametric regression in exponential families. *Ann. Statist.* **38** 2005–2046. MR2676882 https://doi.org/10.1214/09-AOS762

[7] CATTANEO, M. D., CHANDAK, R., JANSSON, M. and MA, X. (2024). Boundary adaptive local polynomial conditional density estimators. *Bernoulli* **30** 3193–3223. MR4779862 https://doi.org/10.3150/23-bej1711

[8] CATTANEO, M. D., FARRELL, M. H. and FENG, Y. (2020). Large sample properties of partitioning-based series estimators. *Ann. Statist.* **48** 1718–1741. MR4124341 https://doi.org/10.1214/19-AOS1865

[9] CATTANEO, M. D., FENG, Y. and UNDERWOOD, W. G. (2024). Uniform inference for kernel density estimators with dyadic data. *J. Amer. Statist. Assoc.* **119** 2695–2708. MR4833908 https://doi.org/10.1080/01621459.2023.2272785

MSC2020 subject classifications. Primary 60F17; secondary 62E17, 62G20.

Key words and phrases. Empirical processes, strong approximation, Gaussian approximation, uniform inference, local empirical process, nonparametric regression.

[10] CATTANEO, M. D., JANSSON, M. and MA, X. (2024). Local regression distribution estimators. *J. Econometrics* **240** Paper No. 105074. MR4719214 https://doi.org/10.1016/j.jeconom.2021.01.006

[11] CATTANEO, M. D., MASINI, R. P. and UNDERWOOD, W. G. (2025). Yurinskii's Coupling for Martingales. *Ann. Statist.*, to appear. Available at arXiv:2210.00362.

[12] CATTANEO, M. D. and YU, R. (2025). Supplement to "Strong approximations for empirical processes indexed by Lipschitz functions." https://doi.org/10.1214/25-AOS2500SUPP

[13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

[14] CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. *Probability and Mathematical Statistics*. Academic Press, New York. MR0666546

[15] DEDECKER, J., MERLEVÈDE, F. and RIO, E. (2014). Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electron. J. Probab.* **19** no. 9. MR3164762 https://doi.org/10.1214/EJP.v19-2658

[16] EINMAHL, U. and MASON, D. M. (1998). Strong approximations to the local empirical process. In *High Dimensional Probability* (*Oberwolfach*, 1996). *Progress in Probability* **43** 75–92. Birkhäuser, Basel. MR1652321

[17] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

[18] GINÉ, E., KOLTCHINSKII, V. and SAKHANENKO, L. (2004). Kernel density estimators: Convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields* **130** 167–198. MR2093761 https://doi.org/10.1007/s00440-004-0339-x

[19] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738

[20] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*, [40]. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

[21] HUANG, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600–1635. MR2012827 https://doi.org/10.1214/aos/1065705120

[22] JURINSKIĬ, V. V. (1977). The error of Gaussian approximation of convolutions. *Theory Probab. Appl.* **22**. 236–247.

[23] KOLTCHINSKII, V. I. (1994). Komlos-Major-Tusnady approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.* **7** 73–118. MR1256392 https://doi.org/10.1007/BF02213361

[24] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[25] LINDVALL, T. (1992). *Lectures on the Coupling Method*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR1180522

[26] MASON, D. M. and VAN ZWET, W. R. (2011). A refinement of the KMT inequality for the uniform empirical process. In *Selected Works of Willem van Zwet* 415–428. Springer, Berlin.

[27] MASON, D. M. and ZHOU, H. H. (2012). Quantile coupling inequalities and their applications. *Probab. Surv.* **9** 439–479. MR3007210 https://doi.org/10.1214/12-PS198

[28] MASSART, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Probab.* **17** 266–291. MR0972785

[29] POLLARD, D. (2002). *A User's Guide to Measure Theoretic Probability*. *Cambridge Series in Statistical and Probabilistic Mathematics* **8**. Cambridge Univ. Press, Cambridge. MR1873379

[30] RIO, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21–45. MR1254823 https://doi.org/10.1007/BF01311347

[31] SAKHANENKO, A. I. (1996). Estimates for the accuracy of constructions on a probability space in the central limit theorem. *Sib. Math. J.* **37** 811–823.

[32] SAKHANENKO, L. (2015). Asymptotics of suprema of weighted Gaussian fields with applications to kernel density estimators. *Theory Probab. Appl.* **59** 415–451. MR3415976 https://doi.org/10.1137/S0040585X97T987211

[33] SETTATI, A. (2009). Gaussian approximation of the empirical process under random entropy conditions. *Stochastic Process. Appl.* **119** 1541–1560. MR2513118 https://doi.org/10.1016/j.spa.2008.08.001

[34] VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes—with Applications to Statistics*, 2nd ed. *Springer Series in Statistics*. Springer, Cham. MR4628026 https://doi.org/10.1007/978-3-031-29040-4

[35] WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. *Monographs on Statistics and Applied Probability* **60**. CRC Press, London. MR1319818 https://doi.org/10.1007/978-1-4899-4493-1

[36] ZAĬTSEV, A. Y. (1986). Estimates for the Lévy-Prokhorov distance in the multidimensional central limit theorem for random vectors with finite exponential moments. *Theory Probab. Appl.* **31** 203–220.
[37] ZAĬTSEV, A. Y. (2013). The accuracy of strong Gaussian approximation for sums of independent random vectors. *Russian Math. Surveys* **68** 721–761.

The Annals of Statistics
2025, Vol. 53, No. 3, 1230–1256
https://doi.org/10.1214/25-AOS2501
© Institute of Mathematical Statistics, 2025

TESTING STATIONARITY AND CHANGE POINT DETECTION IN REINFORCEMENT LEARNING

BY MENGBING LI[1,a], CHENGCHUN SHI[2,c], ZHENKE WU[1,b] AND PIOTR FRYZLEWICZ[2,d]

[1]*Department of Biostatistics, University of Michigan,* [a]*mengbing@umich.edu,* [b]*zhenkewu@umich.edu*
[2]*Department of Statistics, London School of Economics and Political Science,* [c]*c.shi7@lse.ac.uk,* [d]*p.fryzlewicz@lse.ac.uk*

We consider reinforcement learning (RL) in possibly nonstationary environments. Many existing RL algorithms in the literature rely on the stationarity assumption that requires the state transition and reward functions to be constant over time. However, this assumption is restrictive in practice and is likely to be violated in a number of applications, including traffic signal control, robotics and mobile health. In this paper, we develop a model-free test to assess the stationarity of the optimal Q-function based on pre-collected historical data, without additional online data collection. Based on the proposed test, we further develop a change point detection method that can be naturally coupled with existing state-of-the-art RL methods designed in stationary environments for online policy optimization in nonstationary environments. The usefulness of our method is illustrated by theoretical results, simulation studies, and a real data example from the 2018 Intern Health Study. A Python implementation of the proposed procedure is publicly available at https://github.com/limengbinggz/CUSUM-RL.

REFERENCES

AGARWAL, A., JIANG, N., KAKADE, S. M. and SUN, W. (2019). Reinforcement Learning: Theory and Algorithms. CS Dept., UW Seattle, Seattle, WA, USA, Tech. Rep., 32.

ALEGRE, L. N., BAZZAN, A. L. C. and DA SILVA, B. C. (2021). Minimum-delay adaptation in non-stationary reinforcement learning via online high-confidence change-point detection. In *Proceedings of the* 20*th International Conference on Autonomous Agents and MultiAgent Systems*, *AAMAS* '21 97–105. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

ALQUIER, P., DOUKHAN, P. and FAN, X. (2019). Exponential inequalities for nonstationary Markov chains. *Depend. Model.* **7** 150–168. MR3962650 https://doi.org/10.1515/demo-2019-0007

AMINIKHANGHAHI, S. and COOK, D. (2017). A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **51** 339–367.

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. MR3343791 https://doi.org/10.1016/j.jeconom.2015.02.014

BIAN, Z., SHI, C., QI, Z. and WANG, L. (2024). Off-policy evaluation in doubly inhomogeneous environments. *J. Amer. Statist. Assoc.* To appear.

CAZELLES, B., CHAMPAGNE, C. and DUREAU, J. (2018). Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models. *PLoS Comput. Biol.* **14** e1006211.

CHAKRABORTY, B., LABER, E. B. and ZHAO, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69** 714–723. MR3106599 https://doi.org/10.1111/biom.12052

CHAKRABORTY, B., MURPHY, S. and STRECHER, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.* **19** 317–343. MR2757118 https://doi.org/10.1177/0962280209105013

CHEN, B. and HONG, Y. (2012). Testing for the Markov property in time series. *Econometric Theory* **28** 130–178. MR2899216 https://doi.org/10.1017/S0266466611000065

CHEN, E. Y., SONG, R. and JORDAN, M. I. (2024). Reinforcement learning with heterogeneous data: Estimation and inference. *J. Amer. Statist. Assoc.* **119** 3113–3126.

MSC2020 subject classifications. Primary 68T01, 62G05; secondary 62M05.
Key words and phrases. Reinforcement learning, nonstationarity, hypothesis testing, change point detection.

CHEN, J. and JIANG, N. (2019). Information-theoretic considerations in batch reinforcement learning. In *International Conference on Machine Learning* 1042–1051. PMLR.

CHEN, X. and CHRISTENSEN, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *J. Econometrics* **188** 447–465. MR3383220 https://doi.org/10.1016/j.jeconom.2015.03.010

CHEN, X. and CHRISTENSEN, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quant. Econ.* **9** 39–84. MR3789729 https://doi.org/10.3982/QE722

CHEN, X., WANG, Y. and ZHOU, Y. (2020). Dynamic assortment optimization with changing contextual information. *J. Mach. Learn. Res.* **21** Paper No. 216, 44. MR4209502

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

CHEUNG, W. C., SIMCHI-LEVI, D. and ZHU, R. (2020). Reinforcement learning for non-stationary Markov decision processes: The blessing of (more) optimism. In *International Conference on Machine Learning* 1843–1854. PMLR.

CHO, H. and FRYZLEWICZ, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statist. Sinica* **22** 207–229. MR2933173 https://doi.org/10.5705/ss.2009.280

CHO, H. and FRYZLEWICZ, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 https://doi.org/10.1111/rssb.12079

COLLINS, L. M., MURPHY, S. A. and STRECHER, V. (2007). The multiphase optimization strategy (most) and the sequential multiple assignment randomized trial (smart): New methods for more potent ehealth interventions. *Amer. J. Prev. Med.* **32** S112–S118.

COOPERBERG, M. R., BROERING, J. M., LITWIN, M. S., LUBECK, D. P., MEHTA, S. S., HENNING, J. M., CARROLL, P. R. and INVESTIGATORS, C. (2004). The contemporary management of prostate cancer in the United States: Lessons from the cancer of the prostate strategic urologic research endeavor (capsure), a national disease registry. *J. Urol.* **171** 1393–1401.

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. With a foreword by David Kendall. MR2743035

DEDECKER, J. and FAN, X. (2015). Deviation inequalities for separately Lipschitz functionals of iterated random functions. *Stochastic Process. Appl.* **125** 60–90. MR3274692 https://doi.org/10.1016/j.spa.2014.08.001

DOMINGUES, O. D., MÉNARD, P., PIROTTA, M., KAUFMANN, E. and VALKO, M. (2021). A kernel-based approach to non-stationary reinforcement learning in metric spaces. In *International Conference on Artificial Intelligence and Statistics* 3538–3546. PMLR.

EFTEKHARI, H., MUKHERJEE, D., BANERJEE, M. and RITOV, Y. (2020). Markovian snd non-Markovian processes with active decision making strategies for addressing the COVID-19 pandemic. arXiv preprint. Available at arXiv:2008.00375.

EICHENBAUM, M. S., REBELO, S. and TRABANDT, M. (2020). The macroeconomics of epidemics. Technical report, National Bureau of Economic Research.

ERNST, D., GEURTS, P. and WEHENKEL, L. (2005). Tree-based batch mode reinforcement learning. *J. Mach. Learn. Res.* **6** 503–556. MR2249830

ERTEFAIE, A. and STRAWDERMAN, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika* **105** 963–977. MR3877877 https://doi.org/10.1093/biomet/asy043

FAN, J., WANG, Z., XIE, Y. and YANG, Z. (2020). A theoretical analysis of deep q-learning. In *Learning for Dynamics and Control* 486–489. PMLR.

FANG, E. X., WANG, Z. and WANG, L. (2023). Fairness-oriented learning for optimal individualized treatment rules. *J. Amer. Statist. Assoc.* **118** 1733–1746. MR4646602 https://doi.org/10.1080/01621459.2021.2008402

FEI, Y., YANG, Z., WANG, Z. and XIE, Q. (2020). Dynamic regret of policy optimization in non-stationary environments. In *Advances in Neural Information Processing Systems* 6743–6754.

FENG, S., YIN, M., HUANG, R., WANG, Y.-X., YANG, J. and LIANG, Y. (2023). Non-stationary reinforcement learning under general function approximation. In *International Conference on Machine Learning* 9976–10007. PMLR.

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

GRENANDER, U. (1981). *Abstract Inference*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0599175

GUO, X. and HE, X. (2021). Inference on selected subgroups in clinical trials. *J. Amer. Statist. Assoc.* **116** 1498–1506. MR4309288 https://doi.org/10.1080/01621459.2020.1740096

HAO, B., JI, X., DUAN, Y., LU, H., SZEPESVARI, C. and WANG, M. (2021). Bootstrapping fitted q-evaluation for off-policy inference. In *International Conference on Machine Learning* 4074–4084. PMLR.

HASSELT, H. V. (2010). Double Q-learning. In *Advances in Neural Information Processing Systems* 2613–2621.

HU, X., QIAN, M., CHENG, B. and CHEUNG, Y. K. (2021). Personalized policy learning using longitudinal mobile health data. *J. Amer. Statist. Assoc.* **116** 410–420. MR4227703 https://doi.org/10.1080/01621459.2020.1785476

HU, Y. and WAGER, S. (2023). Off-policy evaluation in partially observed Markov decision processes under sequential ignorability. *Ann. Statist.* **51** 1561–1585. MR4658568 https://doi.org/10.1214/23-aos2287

HUANG, J. Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist.* **26** 242–272. MR1611780 https://doi.org/10.1214/aos/1030563984

JUDD, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA. MR1646331

KHARROUBI, S. and SALEH, F. (2020). Are lockdown measures effective against COVID-19? *Front. Public Health* **8** 549692. https://doi.org/10.3389/fpubh.2020.549692

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

KLASNJA, P., SMITH, S., SEEWALD, N. J., LEE, A., HALL, K., LUERS, B., HEKLER, E. B. and MURPHY, S. A. (2019). Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of heartsteps. *Ann. Behav. Med.* **53** 573–582.

KOMPELLA, V., CAPOBIANCO, R., JONG, S., BROWNE, J., FOX, S., MEYERS, L., WURMAN, P. and STONE, P. (2020). Reinforcement learning for optimization of COVID-19 mitigation policies. arXiv preprint. Available at arXiv:2010.10560.

KOSOROK, M. R. and LABER, E. B. (2019). Precision medicine. *Annu. Rev. Stat. Appl.* **6** 263–286. MR3939521 https://doi.org/10.1146/annurev-statistics-030718-105251

LE, H., VOLOSHIN, C. and YUE, Y. (2019). Batch policy learning under constraints. In *International Conference on Machine Learning* 3703–3712.

LECARPENTIER, E. and RACHELSON, E. (2019). Non-stationary Markov decision processes, a worst-case approach using model-based reinforcement learning. *Adv. Neural Inf. Process. Syst.* **32**.

LI, G., SHI, L., CHEN, Y., CHI, Y. and WEI, Y. (2024). Settling the sample complexity of model-based offline reinforcement learning. *Ann. Statist.* **52** 233–260. MR4718414 https://doi.org/10.1214/23-aos2342

LI, M., SHI, C., WU, Z. and FRYZLEWICZ, P. (2025). Supplement to "Testing Stationarity and Change Point Detection in Reinforcement Learning." https://doi.org/10.1214/25-AOS2501SUPPA, https://doi.org/10.1214/25-AOS2501SUPPB

LI, Y. (2019). Reinforcement learning applications. arXiv preprint. Available at arXiv:1908.06973.

LIAO, P., GREENEWALD, K., KLASNJA, P. and MURPHY, S. (2020). Personalized HeartSteps: A reinforcement learning algorithm for optimizing physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **4** 1–22.

LIAO, P., KLASNJA, P. and MURPHY, S. (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *J. Amer. Statist. Assoc.* **116** 382–391. MR4227701 https://doi.org/10.1080/01621459.2020.1807993

LIAO, P., QI, Z., WAN, R., KLASNJA, P. and MURPHY, S. A. (2022). Batch policy learning in average reward Markov decision processes. *Ann. Statist.* **50** 3364–3387. MR4524500 https://doi.org/10.1214/22-aos2231

LIU, W., TU, J., ZHANG, Y. and CHEN, X. (2023). Online estimation and inference for robust policy evaluation in reinforcement learning. arXiv preprint. Available at arXiv:2310.02581.

LOPES-JÚNIOR, L. C., BOMFIM, E., DA SILVEIRA, D. S. C., PESSANHA, R. M., SCHUAB, S. I. P. C. and LIMA, R. A. G. (2020). Protocol: Effectiveness of mass testing for control of covid-19: A systematic review protocol. *BMJ Open* **10**.

LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 https://doi.org/10.1080/01621459.2018.1537919

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. MR3476615 https://doi.org/10.1214/15-AOS1384

MAEI, H. R., SZEPESVÁRI, C., BHATNAGAR, S. and SUTTON, R. S. (2010). Toward off-policy learning control with function approximation. In *ICML* 719–726.

MARLING, C. and BUNESCU, R. (2020). The OhioT1DM dataset for blood glucose level prediction: Update 2020. In *CEUR Workshop Proceedings* **2675** 71.

MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628. MR0358933 https://doi.org/10.1214/aop/1176996608

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584 https://doi.org/10.1198/jasa.2009.tm08647

MNIH, V., KAVUKCUOGLU, K., SILVER, D., RUSU, A. A., VENESS, J., BELLEMARE, M. G., GRAVES, A., RIEDMILLER, M., FIDJELAND, A. K. et al. (2015). Human-level control through deep reinforcement learning. *Nature* **518** 529–533.

MUNOS, R. and SZEPESVÁRI, C. (2008). Finite-time bounds for fitted value iteration. *J. Mach. Learn. Res.* **9** 815–857. MR2417255

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389

MYLES, A. J., FEUDALE, R. N., LIU, Y., WOODY, N. A. and BROWN, S. D. (2004). An introduction to decision tree modeling. *J. Chemom.* **18** 275–285.

NECAMP, T., SEN, S., FRANK, E., WALTON, M. A., IONIDES, E. L., FANG, Y., TEWARI, A. and WU, Z. (2020). Assessing real-time moderation for developing adaptive mobile health interventions for medical interns: Micro-randomized trial. *J. Med. Internet Res.* **22** e15033.

NIE, X., BRUNSKILL, E. and WAGER, S. (2021). Learning when-to-treat policies. *J. Amer. Statist. Assoc.* **116** 392–409. MR4227702 https://doi.org/10.1080/01621459.2020.1831925

NIROUI, F., ZHANG, K., KASHINO, Z. and NEJAT, G. (2019). Deep reinforcement learning robot for search and rescue applications: Exploration in unknown cluttered environments. *IEEE Robot. Autom. Lett.* **4** 610–617.

PADAKANDLA, S., PRABUCHANDRAN, K. J. and BHATNAGAR, S. (2020). Reinforcement learning algorithm for non-stationary environments. *Appl. Intell.* **50** 3590–3606.

PRABUCHANDRAN, K. J., SINGH, N., DAYAMA, P., AGARWAL, A. and PANDIT, V. (2022). Change point detection for compositional multivariate data. *Appl. Intell.* **52** 1930–1955.

PUTERMAN, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.* Wiley, New York. A Wiley-Interscience Publication. MR1270015

QI, Z., LIU, D., FU, H. and LIU, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *J. Amer. Statist. Assoc.* **115** 678–691. MR4107672 https://doi.org/10.1080/01621459.2018.1529597

QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351 https://doi.org/10.1214/10-AOS864

QIAN, T., WALTON, A. E., COLLINS, L. M., KLASNJA, P., LANZA, S. T., NAHUM-SHANI, I., RABBI, M., RUSSELL, M. A., WALTON, M. A. et al. (2022). The microrandomized trial for developing digital interventions: Experimental design and data analysis considerations. *Psychol. Methods.*

RAMPRASAD, P., LI, Y., YANG, Z., WANG, Z., SUN, W. W. and CHENG, G. (2023). Online bootstrap inference for policy evaluation in reinforcement learning. *J. Amer. Statist. Assoc.* **118** 2901–2914. MR4681629 https://doi.org/10.1080/01621459.2022.2096620

ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics. Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

SHI, C., FAN, A., SONG, R. and LU, W. (2018a). High-dimensional A-learning for optimal dynamic treatment regimes. *Ann. Statist.* **46** 925–957. MR3797992 https://doi.org/10.1214/17-AOS1570

SHI, C., LU, W. and SONG, R. (2020). Breaking the curse of nonregularity with subagging—inference of the mean outcome under optimal treatment regimes. *J. Mach. Learn. Res.* **21** Paper No. 176, 67. MR4209462

SHI, C., LUO, S., LE, Y., ZHU, H. and SONG, R. (2024a). Statistically efficient advantage learning for offline reinforcement learning in infinite horizons. *J. Amer. Statist. Assoc.* **119** 232–245. MR4713888 https://doi.org/10.1080/01621459.2022.2106868

SHI, C., QI, Z., WANG, J. and ZHOU, F. (2024b). Value enhancement of reinforcement learning via efficient and robust trust region optimization. *J. Amer. Statist. Assoc.* **119** 2011–2025. MR4797919 https://doi.org/10.1080/01621459.2023.2238942

SHI, C., SONG, R., LU, W. and FU, B. (2018b). Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 681–702. MR3849339 https://doi.org/10.1111/rssb.12273

SHI, C., WAN, R., SONG, R., LU, W. and LENG, L. (2020). Does the Markov decision process fit the data: Testing for the Markov property in sequential decision making. In *International Conference on Machine Learning* 8807–8817. PMLR.

SHI, C., ZHANG, S., LU, W. and SONG, R. (2022). Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 765–793. MR4460575

SILVER, D. L., YANG, Q. and LI, L. (2013). Lifelong machine learning systems: Beyond learning algorithms. In 2013 *AAAI Spring Symposium Series.*

SONG, R., WANG, W., ZENG, D. and KOSOROK, M. R. (2015). Penalized Q-learning for dynamic treatment regimens. *Statist. Sinica* **25** 901–920. MR3409730

SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning*: *An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951

TRUONG, C., OUDRE, L. and VAYATIS, N. (2020). Selective review of offline change point detection methods. *Signal Process*. **167** 107299.

TSIATIS, A. A., DAVIDIAN, M., HOLLOWAY, S. T. and LABER, E. B. (2019). *Dynamic Treatment Regimes*: *Statistical Methods for Precision Medicine*. CRC press, Boca Raton.

UEHARA, M., IMAIZUMI, M., JIANG, N., KALLUS, N., SUN, W. and XIE, T. (2021). Finite sample analysis of minimax offline reinforcement learning: Completeness, fast rates and first-order efficiency. arXiv preprint. Available at arXiv:2102.02981.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. With applications to statistics. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

WALLACE, M. P. and MOODIE, E. E. M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71** 636–644. MR3402599 https://doi.org/10.1111/biom.12306

WAN, R., ZHANG, X. and SONG, R. (2020). Multi-objective reinforcement learning for infectious disease control with application to covid-19 spread. arXiv preprint. Available at arXiv:2009.04607.

WANG, J., QI, Z. and WONG, R. K. W. (2023). Projected state-action balancing weights for offline reinforcement learning. *Ann*. *Statist*. **51** 1639–1665. MR4658571 https://doi.org/10.1214/23-aos2302

WANG, J., SHI, C. and WU, Z. (2023). A robust test for the stationarity assumption in sequential decision making. In *International Conference on Machine Learning* 36355–36379. PMLR.

WANG, L., ZHOU, Y., SONG, R. and SHERWOOD, B. (2018). Quantile-optimal treatment regimes. *J*. *Amer*. *Statist*. *Assoc*. **113** 1243–1254. MR3862354 https://doi.org/10.1080/01621459.2017.1330204

WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

WATKINS, C. J. and DAYAN, P. (1992). Q-learning. *Mach*. *Learn*. **8** 279–292.

WEI, C.-Y. and LUO, H. (2021). Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach. In *Conference on Learning Theory* 4300–4354. PMLR.

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann*. *Statist*. **14** 1261–1295. With discussion and a rejoinder by the author. MR0868303 https://doi.org/10.1214/aos/1176350142

XIE, A., HARRISON, J. and FINN, C. (2021). Deep reinforcement learning amidst continual structured non-stationarity. In *International Conference on Machine Learning* 11393–11403. PMLR.

YU, M. and CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445 https://doi.org/10.1093/biomet/ast014

ZHANG, Y., LABER, E. B., DAVIDIAN, M. and TSIATIS, A. A. (2018). Estimation of optimal treatment regimes using lists. *J*. *Amer*. *Statist*. *Assoc*. **113** 1541–1549. MR3902228 https://doi.org/10.1080/01621459.2017.1345743

ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J*. *Amer*. *Statist*. *Assoc*. **110** 583–598. MR3367249 https://doi.org/10.1080/01621459.2014.937488

ZHONG, H., YANG, Z. and SZEPESVÁRI, Z. W. C. (2021). Optimistic policy optimization is provably efficient in non-stationary MDPs. arXiv preprint. Available at arXiv:2110.08984.

ZHOU, W., ZHU, R. and QU, A. (2024). Estimating optimal infinite horizon dynamic treatment regimes via pT-learning. *J*. *Amer*. *Statist*. *Assoc*. **119** 625–638. MR4713918 https://doi.org/10.1080/01621459.2022.2138760

ZHOU, Y., SHI, C., LI, L. and YAO, Q. (2023). Testing for the Markov property in time series via deep conditional generative learning. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **85** 1204–1222. MR4716172 https://doi.org/10.1093/jrsssb/qkad064

ZHU, R., ZHAO, Y.-Q., CHEN, G., MA, S. and ZHAO, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics* **73** 391–400. MR3665956 https://doi.org/10.1111/biom.12593

The Annals of Statistics
2025, Vol. 53, No. 3, 1257–1274
https://doi.org/10.1214/25-AOS2502
© Institute of Mathematical Statistics, 2025

ADAPTIVE ESTIMATION OF THE $\mathbb{L}_2$-NORM OF A PROBABILITY DENSITY AND RELATED TOPICS I. LOWER BOUNDS

BY G. CLEANTHOUS[1,a], A. G. GEORGIADIS[2,b] AND O. V. LEPSKI[3,c]

[1]*Department of Mathematics and Statistics, Maynooth University,* [a]*galatia.cleanthous@mu.ie*
[2]*School of Computer Science and Statistics, Trinity College of Dublin,* [b]*georgiaa@tcd.ie*
[3]*Institut de Mathématique de Marseille, CNRS, Aix-Marseille Université,* [c]*oleg.lepski@univ-amu.fr*

We deal with the problem of the adaptive estimation of the $\mathbb{L}_2$–norm of a probability density on $\mathbb{R}^d$, $d \geq 1$, from independent observations. The unknown density is assumed to be uniformly bounded and to belong to the union of balls in the isotropic/anisotropic Nikolskii's spaces. We will show that the optimally adaptive estimators over the collection of considered functional classes do no exist. Also, in the framework of an abstract density model we present several generic lower bounds related to the adaptive estimation of an arbitrary functional of a probability density. These results having independent interest have no analogue in the existing literature. In the companion paper (Cleanthous, Georgiadis and Lepski (2024)), we prove that established lower bounds are tight and provide with explicit construction of adaptive estimators of $\mathbb{L}_2$–norm of the density.

REFERENCES

BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** 2524–2535. MR1425965 https://doi.org/10.1214/aos/1032181166

BUTUCEA, C. (2000). Two adaptive rates of convergence in pointwise density estimation. *Math. Methods Statist.* **9** 39–64. MR1772224

CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108 https://doi.org/10.1214/009053605000000147

CLEANTHOUS, G., GEORGIADIS, A. G. and LEPSKI, O. V. (2024). Adaptive estimation of $\mathbb{L}_2$-norm of a probability density and related topics II. Adaptive upper bounds via oracle approach. Preprint.

CLEANTHOUS, G., GEORGIADIS, A. G. and LEPSKI, O. V. (2025). Supplement to "Adaptive estimation of the $\mathbb{L}_2$-norm of a probability density and related topics I. Lower bounds." https://doi.org/10.1214/25-AOS2502SUPP

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323. MR1081043 https://doi.org/10.1016/0885-064X(90)90025-9

EFROMOVICH, S. and LOW, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24** 1106–1125. MR1401840 https://doi.org/10.1214/aos/1032526959

GAYRAUD, G. and TRIBOULEY, K. (1999). Wavelet methods to estimate an integrated quadratic functional: Adaptivity and asymptotic law. *Statist. Probab. Lett.* **44** 109–122. MR1706448 https://doi.org/10.1016/S0167-7152(98)00296-X

GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653 https://doi.org/10.3150/07-BEJ110

GOLDENSHLUGER, A. and LEPSKI, O. V. (2022a). Minimax estimation of norms of a probability density: I. Lower bounds. *Bernoulli* **28** 1120–1154. MR4388932 https://doi.org/10.3150/21-bej1380

GOLDENSHLUGER, A. and LEPSKI, O. V. (2022b). Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures. *Bernoulli* **28** 1155–1178. MR4388933 https://doi.org/10.3150/21-bej1381

MSC2020 subject classifications. 62G05, 62G20.
Key words and phrases. Estimation of functionals, minimax risk, anisotropic Nikolskii's class, minimax adaptive estimation, adaptive rate of convergence.

HAN, Y., JIAO, J. and MUKHERJEE, R. (2020). On estimation of L_r-norms in Gaussian white noise models. *Probab. Theory Related Fields* **177** 1243–1294. MR4126939 https://doi.org/10.1007/s00440-020-00982-x

KERKYACHARIAN, G. and PICARD, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* **24** 485–507. MR1394973 https://doi.org/10.1214/aos/1032894450

KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564. MR2214904 https://doi.org/10.1007/s00440-005-0447-2

KLUTCHNIKOFF, N. (2014). Pointwise adaptive estimation of a multivariate function. *Math. Methods Statist.* **23** 132–150. MR3224636 https://doi.org/10.3103/S1066530714020045

LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

LAURENT, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli* **3** 181–211. MR1466306 https://doi.org/10.2307/3318586

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395

LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006

LEPSKI, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. MR1091202 https://doi.org/10.1137/1135065

LEPSKI, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 682–697. MR1147167 https://doi.org/10.1137/1136085

LEPSKI, O. V. (1992). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatn. Primen.* **37** 468–481. MR1214353 https://doi.org/10.1137/1137095

NEMIROVSKIĬ, A. S. (1990). Necessary conditions for efficient estimation of functionals of a nonparametric signal observed in white noise. *Teor. Veroyatn. Primen.* **35** 83–91. MR1050056 https://doi.org/10.1137/1135009

NIKOL'SKII, S. M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya*, 2nd ed. Nauka, Moscow.

REBELLES, G. (2015). Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli* **21** 1984–2023. MR3378457 https://doi.org/10.3150/14-BEJ633

TCHETGEN, E., LI, L., ROBINS, J. and VAN DER VAART, A. (2008). Minimax estimation of the integral of a power of a density. *Statist. Probab. Lett.* **78** 3307–3311. MR2479495 https://doi.org/10.1016/j.spl.2008.07.001

TRIBOULEY, K. (2000). Adaptive estimation of integrated functionals. *Math. Methods Statist.* **9** 19–38. MR1772223

TSYBAKOV, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. MR1700239 https://doi.org/10.1214/aos/1024691478

The Annals of Statistics
2025, Vol. 53, No. 3, 1275–1297
https://doi.org/10.1214/25-AOS2503
© Institute of Mathematical Statistics, 2025

ADAPTIVE ESTIMATION OF THE $\mathbb{L}_2$-NORM OF A PROBABILITY DENSITY AND RELATED TOPICS II. UPPER BOUNDS VIA THE ORACLE APPROACH

BY G. CLEANTHOUS[1,a], A.G. GEORGIADIS[2,b] AND O.V. LEPSKI[3,c]

[1]*Department of Mathematics and Statistics, Maynooth University,* [a]*galatia.cleanthous@mu.ie*
[2]*School of Computer Science and Statistics, Trinity College of Dublin,* [b]*georgiaa@tcd.ie*
[3]*Institut de Mathématique de Marseille, CNRS, Aix-Marseille Université,* [c]*oleg.lepski@univ-amu.fr*

This is the second part of the research project initiated in Cleanthous, Georgiadis and Lepski (2024a). We deal with the problem of the adaptive estimation of the $\mathbb{L}_2$-norm of a probability density on $\mathbb{R}^d$, $d \geq 1$, from independent observations. The unknown density is assumed to be uniformly bounded by unknown constant and to belong to the union of balls in the isotropic/anisotropic Nikolskii's spaces. In Cleanthous, Georgiadis and Lepski (2024a), we have proved that the optimally adaptive estimators do no exist in the considered problem and provided with several lower bounds for the adaptive risk. In this part, we show that these bounds are tight and present the adaptive estimator which is obtained by a data-driven selection from a family of kernel-based estimators. The proposed estimation procedure as well as the computation of its risk are heavily based on new concentration inequalities for decoupled U-statistics of order two established in Section 4. It is also worth noting that all our results are derived from the unique oracle inequality which may be of independent interest.

REFERENCES

CLEANTHOUS, G., GEORGIADIS, A. G. and LEPSKI, O. V. (2024). Adaptive estimation of $\mathbb{L}_2$-norm of a probability density and related topics I. Lower bounds. Preprint.

CLEANTHOUS, G., GEORGIADIS, A. and LEPSKI, O. (2025). Supplement to "Adaptive estimation of the $\mathbb{L}_2$-norm of a probability density and related topics II. Upper bounds via the oracle approach https://doi.org/10.1214/25-AOS2503SUPP

DE LA PEÑA, V. H. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** 806–816. MR1334173

FOLLAND, G. B. (1999). *Real Analysis*: *Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics* (*New York*). Wiley, New York. MR1681462

GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability*, *II* (*Seattle*, *WA*, 1999). *Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

GOLDENSHLUGER, A. and LEPSKI, O. V. (2022a). Minimax estimation of norms of a probability density: I. Lower bounds. *Bernoulli* **28** 1120–1154. MR4388932 https://doi.org/10.3150/21-bej1380

GOLDENSHLUGER, A. and LEPSKI, O. V. (2022b). Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures. *Bernoulli* **28** 1155–1178. MR4388933 https://doi.org/10.3150/21-bej1381

HOUDRÉ, C. and REYNAUD-BOURET, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 55–69. Birkhäuser, Basel. MR2073426

KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916 https://doi.org/10.1007/PL00008800

LEPSKI, O. V. (2018). A new approach to estimator selection. *Bernoulli* **24** 2776–2810. MR3779702 https://doi.org/10.3150/17-BEJ945

LEPSKI, O. V. (2023). Theory of adaptive estimation. In *ICM—International Congress of Mathematicians*. *Vol*. 7. *Sections* 15–20 5478–5498. EMS Press, Berlin. MR4680451

MSC2020 subject classifications. 62G05, 62G20.
Key words and phrases. $\mathbb{L}_2$-norm estimation, U-statistics, minimax risk, anisotropic Nikolskii's class, minimax adaptive estimation, oracle approach, data-driven selection.

The Annals of Statistics
2025, Vol. 53, No. 3, 1298–1318
https://doi.org/10.1214/25-AOS2504
© Institute of Mathematical Statistics, 2025

ASYMPTOTIC DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATOR IN GENERALIZED LINEAR MIXED MODELS WITH CROSSED RANDOM EFFECTS

BY JIMING JIANG[1,a]

[1] *Department of Statistics, University of California, Davis,* [a] *jimjiang@ucdavis.edu*

Generalized linear mixed models (GLMM) with crossed random effects is infamously known to present major challenges not only computationally but also theoretically. In fact, to date only consistency of the maximum likelihood estimators (MLE) has been proved for GLMM with crosses random effects. We introduce a new technique in asymptotic analysis built on a second-order Laplace approximation (LA) of a conditional expectation, whose coefficients are carefully evaluated. The LA leads to a large system of equations involving the conditional expectations, which is asymptotically inverted to obtain explicit approximation to the conditional expectations. This powerful technique leads to a new approach to establishing asymptotic distribution of the MLE when the likelihood function is intractable. As a result, asymptotic normality of the MLE is rigorously established for GLMM with crossed random effects, bringing answer to an open problem dating back to decades ago.

REFERENCES

BELLIO, R., GHOSH, W., OWEN, A. B. and VARIN, C. (2023). Scalable estimation of probit models with crossed random effects. Available at arXiv:2308.15681.

BOOTH, J. G. and HOBERT, J. P. (1999). Maximum generalized linear mixed model likelihood with an automated Monte Carlo EM algorithm. *J. R. Stat. Soc. B* **61** 265–285.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

EKVALL, K. O. and JONES, G. L. (2020). Consistent maximum likelihood estimation using subsets with applications to multivariate mixed models. *Ann. Statist.* **48** 932–952. MR4102682 https://doi.org/10.1214/19-AOS1830

GHANDWANI, D., GHOSH, S., HASTIE, T. and OWEN, A. B. (2023). Scalable solution to crossed random effects model with random slopes. Available at arXiv:2307:12378.

JIANG, J. (2013). The subset argument and consistency of MLE in GLMM: Answer to an open problem and beyond. *Ann. Statist.* **41** 177–195. MR3059414 https://doi.org/10.1214/13-AOS1084

JIANG, J. (2017). *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems. Monographs on Statistics and Applied Probability* **155**. CRC Press, Boca Raton, FL. MR3676375

JIANG, J. (2022). *Large Sample Techniques for Statistics. Springer Texts in Statistics*. Springer, Cham. Second edition [of 2675055]. MR4417782 https://doi.org/10.1007/978-3-030-91695-4

JIANG, J. (2025). Supplement to "Asymptotic Distribution of Maximum Likelihood Estimator in Generalized Linear Mixed Models with Crossed Random Effects." https://doi.org/10.1214/25-AOS2504SUPP

JIANG, J. and NGUYEN, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications. Springer Series in Statistics*. Springer, New York. Second edition [of 2308058]. MR4269778 https://doi.org/10.1007/978-1-0716-1282-8

JIANG, J., WAND, M. P. and BHASKARAN, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 55–82. MR4400390

JOHNSTONE, I. (2024). Expectation propagation and maximum likelihood in generalized linear mixed models, presentation at *Statistics in the Age of AI Conference, May 9–11, 2024, Washington, D.C.* Available at https://statistics.columbian.gwu.edu/statistics-age-ai.

MSC2020 subject classifications. Primary 62F12; secondary 62J12.

Key words and phrases. Approximation, asymptotic normality, crossed random effects, GLMM, high-dimensional integral, Laplace approximation, large-system inversion, maximum likelihood.

McCullagh, P. (2000). Resampling and exchangeable arrays. *Bernoulli* **6** 285–301. MR1748722 https://doi.org/10.2307/3318577

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. CRC Press, London. Second edition [of MR0727836]. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

Torabi, M. (2012). Likelihood inference in generalized linear mixed models with two components of dispersion using data cloning. *Comput*. *Statist*. *Data Anal*. **56** 4259–4265. MR2957869 https://doi.org/10.1016/j.csda.2012.04.008

The Annals of Statistics
2025, Vol. 53, No. 3, 1319–1346
https://doi.org/10.1214/25-AOS2507
© Institute of Mathematical Statistics, 2025

SELF-NORMALIZED CRAMÉR TYPE MODERATE DEVIATION THEOREM FOR GAUSSIAN APPROXIMATION

BY JINGKUN QIU[1,a], SONG XI CHEN[2,b] AND QI-MAN SHAO[3,c]

[1]*Guanghua School of Management, Peking University,* [a]*jkqiu@stu.pku.edu.cn*

[2]*Department of Statistics and Data Science, Tsinghua University,* [b]*sxchen@tsinghua.edu.cn*

[3]*Department of Statistics and Data Science, Shenzhen International Center for Mathematics, National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology,* [c]*shaoqm@sustech.edu.cn*

Berry–Esseen type bounds for Gaussian approximation of standardized sums have been extensively studied under exponential type moment conditions. In this paper, a Cramér type moderate deviation theorem is established for self-normalized Gaussian approximation under finite moment conditions. More specifically, let $X_1, X_2, \dots, X_n$ be i.i.d. $\mathbb{R}^p$-valued random vectors with zero means. Let $S_{n,j} = \sum_{i=1}^n X_{ij}$ and $V_{n,j}^2 = \sum_{i=1}^n X_{ij}^2$. We show that if the correlation matrix of X_1 is I_p and the third moment of X_1 is finite, then

$$\frac{\mathbb{P}(\max_{1\le j\le p} S_{n,j}/V_{n,j} > x)}{\mathbb{P}(\max_{1\le j\le p} Z_j > x)} \to 1$$

uniformly for $0 \le x \le o(n^{1/6})$ and for all $p \ge 1$, where $Z_1, \dots, Z_p$ are independent standard normal random variables. Similar result is also established for large x when X_1 has a general correlation matrix. The proof is based on a new Cramér type moderate deviation theorem for the minimum of several self-normalized sums. As an application, we propose a high dimensional one-sample t-test that allows for an exponential growth of p without requiring the commonly used sub-Gaussian assumption.

REFERENCES

[1] ALESHKYAVICHENE, A. K. (1983). Multidimensional integral limit theorems for probabilities of large deviations. *Teor. Veroyatn. Primen.* **28** 62–82. MR0691468

[2] AVANESOV, V. and BUZUN, N. (2018). Change-point detection in high-dimensional covariance structure. *Electron. J. Stat.* **12** 3254–3294. MR3861282 https://doi.org/10.1214/18-EJS1484

[3] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and WEI, Y. (2018). Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *Ann. Statist.* **46** 3643–3675. MR3852664 https://doi.org/10.1214/17-AOS1671

[4] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097 https://doi.org/10.1093/biomet/asu056

[5] BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

[6] CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

[7] CHANG, J., ZHENG, C., ZHOU, W.-X. and ZHOU, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

[8] CHEN, L., WANG, W. and WU, W. B. (2022). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* **117** 1951–1963. MR4528482 https://doi.org/10.1080/01621459.2021.1893178

[9] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Partial recovery for top-k ranking: Optimality of MLE and suboptimality of the spectral method. *Ann. Statist.* **50** 1618–1652. MR4441134 https://doi.org/10.1214/21-aos2166

MSC2020 subject classifications. Primary 62E17, 62E20; secondary 62G10.
Key words and phrases. Gaussian approximations, moderate deviations, self-normalized.

[10] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[11] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

[12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. MR4583674 https://doi.org/10.1214/22-aap1870

[13] CHERNOZHUKOV, V., HÄRDLE, W. K., HUANG, C. and WANG, W. (2021). LASSO-driven inference in time and space. *Ann. Statist.* **49** 1702–1735. MR4298878 https://doi.org/10.1214/20-aos2019

[14] CHERNOZHUOKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. MR4500619 https://doi.org/10.1214/22-aos2193

[15] CHETVERIKOV, D., LARSEN, B. and PALMER, C. (2016). IV quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica* **84** 809–833. MR3481380 https://doi.org/10.3982/ECTA12121

[16] CHISTYAKOV, G. P. and GÖTZE, F. (2003). On bounds for moderate deviations for Student's statistic. *Teor. Veroyatn. Primen.* **48** 609–615. MR2141355 https://doi.org/10.1137/S0040585X97980610

[17] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actual. Sci. Ind.* **736** 5–23.

[18] DENG, H. and ZHANG, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

[19] DETTE, H. and GÖSMANN, J. (2018). Relevant change points in high dimensional time series. *Electron. J. Stat.* **12** 2578–2636. MR3849896 https://doi.org/10.1214/18-EJS1464

[20] FAN, J., LOU, Z. and YU, M. (2024). Are latent factor regression and sparse regression adequate? *J. Amer. Statist. Assoc.* **119** 1076–1088. MR4765971 https://doi.org/10.1080/01621459.2023.2169700

[21] FAN, J., MASINI, R. P. and MEDEIROS, M. C. (2023). Bridging factor and sparse models. *Ann. Statist.* **51** 1692–1717. MR4658573 https://doi.org/10.1214/23-aos2304

[22] FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

[23] GÖSMANN, J., STOEHR, C., HEINY, J. and DETTE, H. (2022). Sequential change point detection in high dimensional time series. *Electron. J. Stat.* **16** 3608–3671. MR4444665 https://doi.org/10.1214/22-ejs2027

[24] GUO, X. and TANG, C. Y. (2021). Specification tests for covariance structures in high-dimensional statistical models. *Biometrika* **108** 335–351. MR4259135 https://doi.org/10.1093/biomet/asaa073

[25] HASHORVA, E. and HÜSLER, J. (2003). On multivariate Gaussian tails. *Ann. Inst. Statist. Math.* **55** 507–522. MR2007795 https://doi.org/10.1007/BF02517804

[26] JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616 https://doi.org/10.1214/aop/1068646382

[27] KIM, B., LIU, S. and KOLAR, M. (2021). Two-sample inference for high-dimensional Markov networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 939–962. MR4349123 https://doi.org/10.1111/rssb.12446

[28] KLAASSEN, S., KUECK, J., SPINDLER, M. and CHERNOZHUKOV, V. (2023). Uniform inference in high-dimensional Gaussian graphical models. *Biometrika* **110** 51–68. MR4565443 https://doi.org/10.1093/biomet/asac030

[29] KUCHIBHOTLA, A. K., MUKHERJEE, S. and BANERJEE, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217. MR4177366 https://doi.org/10.3150/20-BEJ1233

[30] KUCHIBHOTLA, A. K. and RINALDO, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate. Available at arXiv:2009.13673.

[31] LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

[32] LI, W. V. and SHAO, Q.-M. (2002). A normal comparison inequality and its applications. *Probab. Theory Related Fields* **122** 494–508. MR1902188 https://doi.org/10.1007/s004400100176

[33] LINNIK, J. V. (1961). On the probability of large deviations for the sums of independent variables. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. II* 289–306. Univ. California Press, Berkeley, CA. MR0137142

[34] LIU, Y. and XIE, J. (2019). Accurate and efficient P-value calculation via Gaussian approximation: A novel Monte-Carlo method. *J. Amer. Statist. Assoc.* **114** 384–392. MR3941262 https://doi.org/10.1080/01621459.2017.1407776

[35] LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

[36] LU, J., KOLAR, M. and LIU, H. (2020). Kernel meets sieve: Post-regularization confidence bands for sparse additive model. *J. Amer. Statist. Assoc.* **115** 2084–2099. MR4189778 https://doi.org/10.1080/01621459.2019.1689984

[37] NEYKOV, M., LU, J. and LIU, H. (2019). Combinatorial inference for graphical models. *Ann. Statist.* **47** 795–827. MR3909951 https://doi.org/10.1214/17-AOS1650

[38] PETROV, V. V. (1954). Generalization of Cramér's limit theorem. *Uspekhi Mat. Nauk* **9** 195–202. MR0065058

[39] QIU, J., CHEN, S. X. and SHAO, Q.-M. (2025). Supplement to "Self-normalized Cramér type moderate deviation theorem for Gaussian approximation." https://doi.org/10.1214/25-AOS2507SUPP

[40] RINALDO, A., WASSERMAN, L. and G'SELL, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748 https://doi.org/10.1214/18-AOS1784

[41] RUDZKIS, R. and BAKSHAEV, A. (2017). General theorems on large deviations for random vectors. *Lith. Math. J.* **57** 367–390. MR3685166 https://doi.org/10.1007/s10986-017-9367-y

[42] SAULIS, L. and STATULEVIČIUS, V. A. (1991). *Limit Theorems for Large Deviations*. *Mathematics and Its Applications* (*Soviet Series*) **73**. Kluwer Academic, Dordrecht. Translated and revised from the 1989 Russian original. MR1171883 https://doi.org/10.1007/978-94-011-3530-6

[43] SAULIS, L. I. (1987). General lemmas on large deviations for a random vector with regular cumulant behavior. I. *Liet. Mat. Rink.* **27** 535–549. MR0925358

[44] SAULIS, L. I. (1987). General lemmas on large deviations for a random vector with regular cumulant behavior. II. *Liet. Mat. Rink.* **27** 747–758. MR0928543

[45] SAULIS, L. I. (1988). General lemmas on large deviations for a random vector with regular cumulant behavior. III. *Liet. Mat. Rink.* **28** 99–111. MR0949646 https://doi.org/10.1007/BF00972252

[46] SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

[47] SHANG, Z. and CHENG, G. (2015). Nonparametric inference in generalized functional linear models. *Ann. Statist.* **43** 1742–1773. MR3357877 https://doi.org/10.1214/15-AOS1322

[48] SHAO, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.* **12** 385–398. MR1684750 https://doi.org/10.1023/A:1021626127372

[49] SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022 https://doi.org/10.3150/15-BEJ719

[50] SHI, C. and LI, L. (2022). Testing mediation effects using logic of Boolean matrices. *J. Amer. Statist. Assoc.* **117** 2014–2027. MR4528486 https://doi.org/10.1080/01621459.2021.1895177

[51] VON BAHR, B. (1967). Multi-dimensional integral limit theorems for large deviations. *Ark. Mat.* **7** 89–99. MR0224129 https://doi.org/10.1007/BF02591679

[52] VON BAHR, B. (1967). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88. MR0228044 https://doi.org/10.1007/BF02591678

[53] WANG, G. and FENG, L. (2023). Computationally efficient and data-adaptive changepoint inference in high dimension. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 936–958. MR4735036 https://doi.org/10.1093/jrsssb/qkad048

[54] WANG, G., WEI, Y. and QIAO, S. (2018). *Generalized Inverses*: *Theory and Computations*, 2nd ed. *Developments in Mathematics* **53**. Springer, Singapore. MR3793648 https://doi.org/10.1007/978-981-13-0146-9

[55] XUE, K. and YAO, F. (2020). Distribution and correlation-free two-sample test of high-dimensional means. *Ann. Statist.* **48** 1304–1328. MR4124324 https://doi.org/10.1214/19-AOS1848

[56] YU, M. and CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

[57] YU, X., LI, D. and XUE, L. (2024). Fisher's combined probability test for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **119** 511–524. MR4713910 https://doi.org/10.1080/01621459.2022.2126781

[58] ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

[59] ZHANG, Y. and POLITIS, D. N. (2022). Ridge regression revisited: Debiasing, thresholding and bootstrap. *Ann. Statist.* **50** 1401–1422. MR4441125 https://doi.org/10.1214/21-aos2156

The Annals of Statistics
2025, Vol. 53, No. 3, 1347–1369
https://doi.org/10.1214/25-AOS2509
© Institute of Mathematical Statistics, 2025

SEMIPARAMETRIC ADAPTIVE ESTIMATION UNDER INFORMATIVE SAMPLING

BY KOSUKE MORIKAWA[1,3,a], YOSHIKAZU TERADA[1,2,b] AND JAE KWANG KIM[3,c]

[1]*Graduate School of Engineering Science, Osaka University,* [a]*morikawa@iastate.edu*
[2]*Center for Advanced Integrated Intelligence Research, RIKEN,* [b]*yoshikazu.terada.es@osaka-u.ac.jp*
[3]*Department of Statistics, Iowa State University,* [c]*jkim@iastate.edu*

In probability sampling, sampling weights are often used to remove selection bias in the sample. The Horvitz–Thompson estimator is well known to be consistent and asymptotically normally distributed; however, it is not necessarily efficient. This study derives the semiparametric efficiency bound for various target parameters by considering the survey weights as random variables and consequently proposes two semiparametric estimators with working models on the survey weights. One estimator assumes a reasonable parametric working model, but the other estimator does not require specific working models by using the debiased/double machine learning method. The proposed estimators are consistent, asymptotically normal, and efficient in a class of regular and asymptotically linear estimators. A limited simulation study is conducted to investigate the finite sample performance of the proposed method. The proposed method is applied to the 1999 Canadian Workplace and Employee Survey data.

REFERENCES

[1] BEAUMONT, J.-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika* **95** 539–553. MR2443174 https://doi.org/10.1093/biomet/asn028
[2] BICKEL, P. J., KLAASSEN, C. A. J. and RITOV, Y. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.
[3] BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144 https://doi.org/10.2307/1402588
[4] CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. MR2335249 https://doi.org/10.1007/s10208-006-0196-8
[5] CHAMBERS, R. L. and SKINNER, C. J. E. . (2003). *Analysis of Survey Data*. Wiley, New York.
[6] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097
[7] FERRARI, S. L. P. and CRIBARI-NETO, F. (2004). Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31** 799–815. MR2095753 https://doi.org/10.1080/0266476042000214501
[8] FULLER, W. A. (2009). *Sampling Statistics*. Wiley, Hoboken, NJ.
[9] GODAMBE, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63** 277–284. MR0428527 https://doi.org/10.1093/biomet/63.2.277
[10] HAN, Q. and WELLNER, J. A. (2021). Complex sampling designs: Uniform limit theorems and applications. *Ann. Statist.* **49** 459–485. MR4206686 https://doi.org/10.1214/20-AOS1964
[11] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7
[12] HEERINGA, S. G., WEST, B. T. and BERGLUND, P. A. (2010). *Applied Survey Data Analysis*. CRC Press/CRC Press, Boca Raton, FL.
[13] ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029

MSC2020 subject classifications. Primary 62F35, 62G20; secondary 62G10.
Key words and phrases. Double/debiased machine learning, semiparametric efficency bound, weight model.

[14] KIM, J. K. and SKINNER, C. J. (2013). Weighting in survey analysis under informative sampling. *Biometrika* **100** 385–398. MR3068441 https://doi.org/10.1093/biomet/ass085

[15] KORN, E. L. and GRAUBARD, B. I. (1999). *Analysis of Health Surveys*. Wiley, New York.

[16] LI, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* **98** 355–370. MR2806433 https://doi.org/10.1093/biomet/asq080

[17] MA, Y. and CARROLL, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *J. Amer. Statist. Assoc.* **101** 1465–1474. MR2279472 https://doi.org/10.1198/016214506000000519

[18] MAGEE, L. (1998). Improving survey-weighted least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 115–126. MR1625652 https://doi.org/10.1111/1467-9868.00112

[19] MCNENEY, B. and WELLNER, J. A. (2000). Application of convolution theorems in semiparametric models with non-i.i.d. data. *J. Statist. Plann. Inference* **91** 441–480.

[20] MORIKAWA, K. and KIM, J. K. (2021). Semiparametric optimal estimation with nonignorable nonresponse data. *Ann. Statist.* **49** 2991–3014. MR4338901 https://doi.org/10.1214/21-aos2070

[21] NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, *Vol. IV* (R. F. Engle and D. L. McFadden, eds.). *Handbooks in Econom.* **2** 2111–2245. North-Holland, Amsterdam. MR1315971

[22] PATAK, Z., HIDIROGLOU, M. and LAVALLÉE, P. (1998). The methodology of the workplace employee survey. In *Proc. Survey Res. Meth. Sect., Am. Statist. Ass.* 83–91. Amer. Statist. Assoc., Washington DC.

[23] PFEFFERMANN, D. and SVERCHKOV, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya, Ser. B* **61** 166–186. MR1720710

[24] PFEFFERMANN, D. and SVERCHKOV, M. Y. (2009). Inference under informative sampling. In *Handbook of Statistics* (D. Pfeffermann and C. R. Rao, eds.) Elsevier, Amsterdam, North Holland.

[25] REDNER, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Ann. Statist.* **9** 225–228. MR0600553

[26] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

[27] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90** 106–121. MR1325118

[28] ROTNITZKY, A. and ROBINS, J. M. (1997). Analysis of semi-parametric regression models with non-ignorable non-response. *Stat. Med.* **16** 81–102.

[29] ROTNITZKY, A., ROBINS, J. M. and SCHARFSTEIN, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Amer. Statist. Assoc.* **93** 1321–1339. MR1666631 https://doi.org/10.2307/2670049

[30] SCHÖLKOPF, B. and SMOLA, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA.

[31] TAN, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97** 661–682. MR2672490 https://doi.org/10.1093/biomet/asq035

[32] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data. Springer Series in Statistics*. Springer, New York. MR2233926

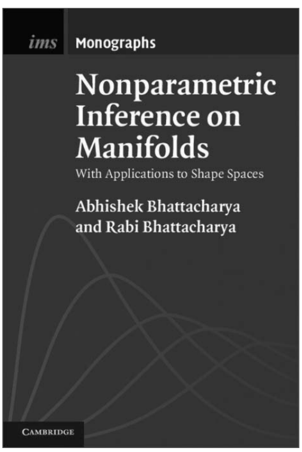
[33] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[34] YUAN, K.-H. and JENNRICH, R. I. (1998). Asymptotics of estimating equations under natural conditions. *J. Multivariate Anal.* **65** 245–260. MR1625893 https://doi.org/10.1006/jmva.1997.1731

[35] ZHAO, J. and MA, Y. (2022). A versatile estimation procedure without estimating the nonignorable missingness mechanism. *J. Amer. Statist. Assoc.* **117** 1916–1930. MR4528480 https://doi.org/10.1080/01621459.2021.1893176

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Nonparametric Inference on Manifolds
With Applications to Shape Spaces

Abhishek Bhattacharya, Rabi Bhattacharya

This book introduces in a systematic manner a general nonparametric theory of statistics on manifolds, with emphasis on manifolds of shapes. The theory has important and varied applications in medical diagnostics, image analysis, and machine vision. An early chapter of examples establishes the effectiveness of the new methods and demonstrates how they outperform their parametric counterparts. Inference is developed for both intrinsic and extrinsic Fréchet means of probability distributions on manifolds, then applied to shape spaces defined as orbits of landmarks under a Lie group of transformations—in particular, similarity, reflection similarity, affine and projective transformations. In addition, nonparametric Bayesian theory is adapted and extended to manifolds for the purposes of density estimation, regression and classification. Ideal for statisticians who analyze manifold data and wish to develop their own methodology, this book is also of interest to probabilists, mathematicians, computer scientists and morphometricians with mathematical training.

IMS member? Claim your 40% discount:
www.cambridge.org/ims

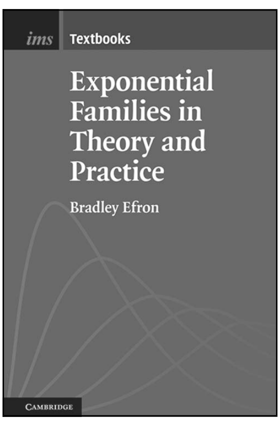
Hardback price US$51.00
(non-member price $85.00)

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.

NEW BOOK

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS



Exponential Families in Theory and Practice

Bradley Efron, Stanford University

During the past half-century, exponential families have attained a position at the center of parametric statistical inference. Theoretical advances have been matched, and more than matched, in the world of applications, where logistic regression by itself has become the go-to methodology in medical statistics, computer-based prediction algorithms, and the social sciences. This book is based on a one-semester graduate course for first year Ph.D. and advanced master's students. After presenting the basic structure of univariate and multivariate exponential families, their application to generalized linear models including logistic and Poisson regression is described in detail, emphasizing geometrical ideas, computational practice, and the analogy with ordinary linear regression. Connections are made with a variety of current statistical methodologies: missing data, survival analysis and proportional hazards, false discovery rates, bootstrapping, and empirical Bayes analysis. The book connects exponential family theory with its applications in a way that doesn't require advanced mathematical preparation.

Hardback $105.00

Paperback $39.99

IMS members are entitled to a 40% discount: email ims@imstat.org to request your code

www.imstat.org/cup/

Cambridge University Press, with the Institute of Mathematical Statistics, established the *IMS Monographs* and *IMS Textbooks* series of high-quality books. The series editors are Mark Handcock, Ramon van Handel, Arnaud Doucet, and John Aston.